SUPPLEMENTO
AL VOLUME XX, SERIE X, DEL
NUOVO CIMENTO
A CURA DELLA SOCIETÀ ITALIANA DI FISICA

1961 2° Trimestre N. 1

An Introduction to the General Theory of Relativity.

F. R. TANGHERLINI (*)

Scuola di Perfezionamento in Fisica Teorica e Nucleare - Napoli

(ricevuto il 5 Agosto 1960)

The purpose of these lectures is to present a brief but comprehensive treatment of the basic physical principles and mathematical techniques employed in the general theory of relativity.

(*) Present address: Departement of Physics, Duke University, Durham, North Carolina.

We shall assume familiarity with the concepts and techniques of special relativity, although for pedagogical purposes we shall review some of the fundamental ideas of special relativity in a way that will serve as an introduction to working with metric tensors different from the diagonal metric tensor used in special relativity.

Frequently the question is raised whether the general theory of relativity is really necessary to handle the problems of gravitation, that is, could not these problems be solved using the « flat » space-time structure of special relativity and interactions analogous to the electromagnetic field? As we shall show more fully in the text, one encounters serious difficulties if one attempts to do this. This is because the electromagnetic field intensities $F_{\mu\nu}$ transform as a tensor, so that we cannot by a co-ordinate transformation, transform away the « intensities » of the electromagnetic field, although to be sure, we can transform to zero some of the components of the $F_{\mu\nu}$ (in general, at a point, how many can we transform to zero?). On the other hand, as is well known, we can transform away locally the intensities of the gravitational field by « freely falling » in it. Hence the components of the gravitational « intensity » do not form a tensor; clearly we need to introduce some new mathematical notions to handle this situation, such as are involved in the Christoffel symbols and covariant differentiation. Thus, although we shall show by explicit calculation that a special relativity type of theory involving a scalar field, does not give the correct answer for the precession of the perihelion of Mercury, we would like to make it clear that this was to be expected from the beginning.

Needless to say, one can try to make more complicated theories based on special relativity, and some of these theories apparently give the correct answer to the famous three checks of general relativity. However, these theories invariably do not make it clear why the gravitational intensities can be transformed away locally, whereas, once one understands the formalism of general relativity the result is obvious.

Having transformed away the intensities of the gravitational field locally, the laws of special relativity once again are valid. The situation is very analogous to that occurring in Riemannian geometry. For example, consider a triangle on the surface of a sphere (say the earth), we know that for sufficiently small triangles Pythagoras' rule holds, but for large triangles this is not the case: the geometry is non-Euclidean. As we shall see general relativity incorporates this idea in such a way that it becomes a cardinal feature of the theory, so that, in general, the geometry will be non-Euclidean, and hence the paths followed by light rays will not be Euclidean straight lines, but their generalization, the so-called null-geodesics. It is mathematically for this reason that light rays are deflected by the presence of a gravitational field, for example, light rays passing near the sun. From a physical standpoint we would say that a light ray, since it possesses energy $E = h\nu$, by the principle of equi-

352

valence also has mass $m = E/c^2 = h\nu/c^2$ and hence « falls » in the gravitational field, or is attracted by it and this gives rise to the deflection. However, a straightforward application of Newton's laws only gives $\frac{1}{2}$ the correct answer. The solution is provided by general relativity which also takes into account the non-Euclidean structure of the space and provides the additional deflection.

Needless to say, once again, one might try to maintain special relativity and argue that the light rays get scattered by the gravitational field; for example, as they do in quantum electrodynamics by a Coulomb field (Delbrück scattering). However, this type of process is extremely complicated and is essentially non-linear so that we do not simplify matters by proceeding in this way. Moreover, from a logical standpoint it would not be satisfactory —because one would still have the line element of special relativity which implies the null geodesics of the space are those traversed by light rays, which would no longer be the case in this type of process.

It is hoped the above discussion makes it clear that the underlying logical construction of the general theory arises to fit the peculiar needs posed by the problems involved in gravitation, and to do so in as systematic and simple a way as possible.

No pretext is made that the theory is in any sense final. The present lack of a unified field theory makes it clear that it cannot be, and indeed it was towards rectifying this lacuna that Einstein devoted nearly forty years of his life.

The enormous difficulties posed by this problem, together with the relatively few checks of general relativity and the growth of more fertile fields of investigation such as atomic and nuclear physics, elementary particles, etc., led to a decline in interest in the general theory for many years. However, at the present time the theory is meeting renewed interest: this is partly due to the development of new experimental techniques and also the conjecture on the part of some theoreticians that the fundamental difficulties that confront quantum field theory may find their resolution in suitably combining the two disciplines. However, since the gravitational potential energy between two electrons 10^{-13} cm apart is $Gm_e^2/(10^{-13}\text{ cm}) \approx 10^{-43}$ MeV, it is clear that only a subtle application of the theory will suffice to give effects at this distance comparable with those of other forces, for examples, the electromagnetic potential energy for the same situation is ≈ 0.5 MeV.

An important application of the general theory is in the field of cosmology. This follows from the fact that the overall space-time structure is intimately related to the gravitational field via the metric tensor. Moreover, according to Mach's principle, the distant masses of the universe are responsible for the inertial effects experienced locally. So that a general theory of relativity must of necessity have a great deal to say about cosmology. We shall try to describe some of these problems here as well as a few simple cosmological models.

CHAPTER I.

SPECIAL RELATIVITY WITH A NON-DIAGONAL LINE ELEMENT

1·1. – How can we measure the velocity of light?

As a method of gaining familiarity with metric tensors different from the one employed in special relativity, $\eta_{\mu\nu}$ = diagonal tensor $(1, -1, -1, -1)$, we shall present an alternative derivation of the Lorentz transformation, which will also serve as a review of the basic ideas of special relativity and an introduction to the concepts and techniques of the general theory.

It is convenient to begin by asking ourselves the question, « How can one measure the velocity of light? » We may assume that there are objects called « clocks » and « rigid rods » available, and that we can measure a distance between two points A and B in our space, by simply sighting along the light path from A to B and along this path introduce a series of unit rods layed end to end, and thereby determine the distance from A to B.

(Clearly in such a method we are assuming the straight line distance between A and B lies along the light path. Another method would be to tie a wire or string to a fixed object at A and tie it also at B and tighten it, and then lay off our distance along this wire with our unit rods. Clearly this assumes we have sufficiently inert bodies at A and B to « anchor » our wire and also that in tightening the wire it will assume the least distance between A and B which we may define to be the straight line distance between A and B. We present these considerations to bring out the important point that concepts such as « distance » have no physical meaning unless we provide an operational definition of how one goes about determining them. For example the well-know definition of the meter as the length of a bar of platinum-iridium kept at Sèvres, is not of much help unless we can transfer the equivalent of such a length to other places and also have a well-defined operation of addition. Note that we can measure fractions of the meter using a « cut and try » process.)

Assuming we have determined the distance from A to B by some such process, as above, and assuming we have two similar clocks one located at A and one located at B, how would we determine the speed of light from A to B? We may imagine a device which sends out a light signal from A and trips the clock at A, so that the time t_A that the light signal left A is recorded, and also a similar device for receiving the light signal at B, so that the time t_B of arrival is recorded. However, this information does not help us at all unless we know also the relative synchronization of the clock at B with the clock at A. How do we determine this synchronization? For, without knowing it,

354

it is quite *impossible* to measure the time of flight from A to B of the light signal or indeed any object. This constitutes one of the many fundamental observations made by EINSTEIN in formulating the special theory of relativity, and it is basic to any discussion of relativity.

1.2. – Synchronization methods.

1.2.1. *Out-and-back method.* – Various methods for synchronization and determining the speed present themselves. One method would be to send the light signal from A to B where it is reflected by a mirror and returned to A, tripping the clock at the time t_A upon leaving and the time t'_A upon returning, the round trips speed of light v_{RT} might then be defined as

$$v_{\mathrm{RT}} = \frac{AB + BA}{t'_A - t_A}, \tag{1.1}$$

where $\overline{AB}$ = distance from A to B, and since we shall assume our geometry is such that $\overline{AB} = \overline{BA}$, we have $v_{\mathrm{RT}} = 2\overline{AB}/(t'_A - t_A)$. However, from this result, can we infer what was the speed of light from A to B, and from B to A? Clearly the method says nothing about how long it took light to travel from A to B and it would be an assumption on our part that this time interval is given by $\Delta t_{AB} = \overline{AB}/v_{\mathrm{RT}}$. Nevertheless, this is the assumption that is made in special relativity, so that a clock at B is said to be in synchronization with a clock at A, if a light signal having left the clock at time t_A arrives at B at a time t_B given by

$$t_B = t_A + \overline{AB}/v_{\mathrm{RT}}\,. \tag{1.2}$$

The quantity v_{RT} is customarily denoted by « c ».

Now, it is fairly clear, such an assumption if it holds in one reference frame which we denote by S and co-ordinates (t, x, y, z), cannot also hold in another frame moving with speed v (as determined by clocks synchronized in the above manner) relative to it, without some dramatic changes in our customary views of space and time.

Before discussing this, however, let us observe that another method of synchronization suggests itself. Thus why not simply imagine two clocks at A synchronized together and move one of them from A to B and compare its time with the clock at B? (We assume that the time of two clocks when they are in the neighborhood of a point can be compared with a delay as small as we please.) Then clock B synchronizes with the clock at A if it synchronizes with the clock that has been moved from A to B. But how do we know nothing happens to the clock when we move it? Let us perform the following « ex-

periment ». We set up our two clocks at A and our light signaling source, and also a receiver at B together with a clock at B. We synchronize B with A by moving the clock from A to B with various speeds (still unknown) and then measure the time for light to travel between A and B using clocks which have been synchronized in this manner. Moreover, we may imagine that in one case we have walked with the clock from A to B, in another case, we have transported the clock from A to B with an airplane, and on still a third occasion we have « shot » the clock from A to B by means of rockets. Then we will find that in the limit as we transported our clock from A to B very slowly, the speed of light determined in the above fashion will approach c. Moreover, we shall find that this result holds in all inertial frames (*).

Since the orientation of the points A and B was unspecified this speed c so determined will clearly be independent of orientation. It also follows that the time for a light signal to go out from A to B and then return will be $2\overline{AB}/c$. So that we see synchronizing the clocks in this way gives us the familiar results. In Sect. 1'3 we shall try to incorporate these ideas into a mathematical scheme, such that if they are true in one inertial frame S they are true in another inertial frame travelling with speed v relative to S.

1'2.2. *Instantaneous synchronization.* – Finally, we should remark still another method of synchronization might have been considered above, namely: synchronization with instantaneous or faster-than-light signals. However, we know experimentally that the speed of light represents an upper limit to the speed with which a signal can be sent between two points. It is a fundamental feature of relativity that this should be the case, so that the speed of light plays the role of a limiting speed. Although of course, mathematically, there is no reason why we cannot consider and calculate effects with such a synchronization procedure. For example, one can show that such signals would lead to causality violations (A. EINSTEIN: *Ann. d. Phys.*, **23**, 371 (1907); see also C. MØLLER, p. 52, W. PAULI, p. 16) if they propagated in a Lorentz-invariant way. If they did not propagate in a Lorentz-invariant way, they would single out *locally* a privileged frame of reference (so-called absolute or ether frame); it would then be very difficult to understand why Lorentz invariance has proven to be so useful and experimentally verifiable over an enormous range of energies.

(*) However, due to the earth's gravitational field and rotation, a co-ordinate system attached to the earth does not constitute an inertial frame. Clearly a better approximation is a freely falling elevator. For simplicity we shall assume we are in outer space where we can neglect permanent gravitational fields and that in fact our co-ordinate system extends to infinity. Our inertial frames will then constitute those frames in uniform motion relative to one another.

1'3. – The line element.

Les us now assume we have synchronized clocks using (1'2.1) or (1'2.2) in an inertial frame S, so that the time it takes light to propagate a distance $\overline{AB}$, is given by

$$\Delta t = c^{-1}\Delta\sigma = c^{-1}(\Delta x^2 + \Delta y^2 + \Delta z^2)^{\frac{1}{2}}\,, \tag{1.3}$$

where Δx, Δy, Δz are the difference in co-ordinates between A and B, and Δt, the time difference as measured by the clocks at A and B. Squaring (1.3) and writing it as

$$c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = 0\,, \tag{1.4}$$

we have the quadratic form characterizing the propagation of light signals. Passing to the limit $\Delta t \to \mathrm{d}t$, $\Delta x \to \mathrm{d}x$, etc., and denoting the left-hand side of (1.4) by $\mathrm{d}s^2$ we arrive at the usual expression for the line element

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2\,. \tag{1.5}$$

The quantity $\mathrm{d}\tau = \mathrm{d}s/c$ is called the « proper time. » The expression (1.5) represents a kind of generalization of the Pythagorean formula; it describes the infinitesimal « distance » between two events. Note that there is zero distance between the two events consisting of a light signal leaving the point A with co-ordinates x_A, y_A, z_A and at time t_A and arriving at B (t_B, x_B, y_B, z_B). That is, for a light signal, $\mathrm{d}s^2 = 0$. We see also if a particle travels with velocity v from a point A to a point B, then since $\mathrm{d}x = v_x\,\mathrm{d}t$, $\mathrm{d}y = v_y\,\mathrm{d}t$ etc., we have

$$\mathrm{d}s^2 = (c^2 - v^2)\,\mathrm{d}t^2\,, \tag{1.6}$$

where $v^2 = v_x^2 + v_y^2 + v_z^2$. Note that $\mathrm{d}s^2 \neq 0$ for such a path unless $v = c$. We see also that for $v > c$, $\mathrm{d}s$ becomes imaginary.

It is now convenient to choose a system of units in which $c = 1$, so that we measure all velocities in units v/c.

As is well known if we now consider another frame S' in motion relative to S with velocity v and look for a co-ordinate transformation connecting them which preserves the same form for the line element, we arrive at the Lorentz transformation. Let us suppose, however, that we did not perform the Lorentz transformation but a Galilean transformation, what form would $\mathrm{d}s^2$ take? (We shall assume the velocity is in the x direction.) We find under the co-

ordinate transformation

$$\left\{\begin{array}{ll} x' = x - vt & y' = y \\ t' = t & z' = z \end{array}\right. \tag{1.7}$$

the line element becomes

$$ds^2 = (1 - v^2)\,dt^2 - 2v\,dt'\,dx' - dx'^2 - dy'^2 - dz'^2 \tag{1.7}$$

and setting $ds^2 = 0$, we have for the propagation time over the path $\Delta x'$, $\Delta y'$, $\Delta z'$,

$$\Delta t' = \frac{v\Delta x'}{1 - v^2} \pm \frac{1}{(1 - v^2)^{\frac{1}{2}}}\left[\frac{\Delta x'^2}{1 - v^2} + \Delta y'^2 + \Delta z'^2\right]^{\frac{1}{2}}, \tag{1.9}$$

in contrast with (1.3). We see from this expression that the propagation time of light is not independent of direction or of the velocity of the frame. There is therefore a fundamental asymmetry between the two frames S' and S. In particular if we send a light signal around any closed path we have:

$$\oint dt' = \frac{1}{(1 - v^2)^{\frac{1}{2}}}\oint\left[\frac{dx'^2}{(1 - v^2)} + dy'^2 + dz'^2\right]^{\frac{1}{2}}, \tag{1.10}$$

since $(v/(1 - v^2))\oint dx'$ vanishes around a closed path. We know, however, experimentally that *no such effects* depending on the velocity of an inertial frame relative to another inertial frame have ever been measured. It follows that the co-ordinate transformation above does not describe what one physically measures in the frame S'. What assumptions were involved in the Galilean transformation? Clearly it assumes in addition to the linearity of the transformation (which is reasonable on physical grounds) that the rods and clocks in S' maintain the same length and interval as seen by an observer in S, and also two events that were simultaneous in S, *i.e.* $\Delta t = 0$ were simultaneous in S', $\Delta t' = 0$ since $\Delta t' = \Delta t$. This latter assumption is reasonable from the standopoint of Newtonian mechanics, since we could in principle synchronize clocks instantaneously, there being no limiting velocity in Newtonian mechanics. On the other hand, it is untenable in relativity as the velocity of light plays the role of a limiting velocity.

To arrive therefore at the « correct » transformation we postulate the following which is fundamental to general relativity:

1) A clock at rest in the primed frame does not read co-ordinate time t', but the *proper time* (interval)

$$ds = (1 - v^2)^{\frac{1}{2}}\,dt'\,. \tag{1.11}$$

It follows from (1.11) that our transformation should therefore introduce a new time given by

$$T = (1 - v^2)^{\frac{1}{2}} t' = (1 - v^2)^{\frac{1}{2}} t . \tag{1.12}$$

The line element now takes the form

$$ds^2 = dT^2 - \frac{2v\,dx'dT}{(1 - v^2)^{\frac{1}{2}}} - dx'^2 - dy'^2 - dz'^2 , \tag{1.13}$$

and (1.9) becomes

$$\Delta T = \frac{v\,\Delta x'}{(1 - v^2)^{\frac{1}{2}}} + \left[\frac{\Delta x'^2}{1 - v^2} + \Delta y'^2 + \Delta z'^2\right]^{\frac{1}{2}} . \tag{1.14}$$

We next postulate:

2) If we measure distance by sending a light signal out and back over a given spatial interval, the elapsed proper time is simply proportional to the euclidean distance independently of the motion of the frame. Then we must have $\Delta T = 2\,\Delta\sigma'$ and this will be the case provided in (1.14) we assume that the length of a rod in S' is given by

$$\Delta X = (1 - v^2)^{-\frac{1}{2}}\,\Delta x' . \tag{1.15}$$

We have therefore by the above assumptions the following co-ordinate transformation

$$\left\{\begin{array}{ll} X = \gamma(x - vt) & Y = y' \\ T = \gamma^{-1} t & Z = z' \end{array}\right. \qquad \text{with} \quad \gamma \equiv (1 - v^2)^{-\frac{1}{2}}, \tag{1.16}$$

and under this co-ordinate transformation the line element takes the form

$$ds^2 = dT^2 - 2v\,dX\,dT - (1 - v^2)\,dX^2 - dY^2 - dZ^2 . \tag{1.17}$$

Now this line element and the corresponding co-ordinate system have a great many properties of the Lorentz transformation, with the exception that two events which are simultaneous in $S(t, x, y, z)$ are also simultaneous in $S'(T, X, Y, Z)$, since $\Delta T = \gamma^{-1}\,\Delta t = 0$ if $\Delta t = 0$. In particular if we send light around any closed path, the time it takes is simply proportional to the length of the path, since

$$\Delta T = v\,\Delta X + \Delta\sigma , \tag{1.18}$$

where $\Delta\sigma = (\Delta X^2 + \Delta Y^2 + \Delta Z^2)^{\frac{1}{2}}$ and

$$\oint \Delta T = \oint v \Delta X + \oint \Delta\sigma = \oint \Delta\sigma \tag{1.19}$$

since $\oint v\,\Delta X = 0$.

However, the clocks have not been synchronized in the appropriate way but rather they have been synchronized absolutely. As a consequence the time taken for light to travel a distance ΔX from A to B is different than that taken from B to A. In fact we see that the time taken from A to B is

$$\frac{\Delta T}{\Delta X} = 1 + v\,, \tag{1.20}$$

and on the return trip

$$\frac{\Delta T}{\Delta X} = 1 - v\,. \tag{1.21}$$

But the average for the two trips out and back is

$$\frac{1}{2}\left(\frac{\Delta T}{\Delta X}\,(\text{out}) + \frac{\Delta T}{\Delta X}\,(\text{back})\right) = 1\,, \tag{1.22}$$

or restoring units c^{-1}, and similarly for the other directions.

However, as we remarked earlier, such a synchronization cannot be achieved without either employing instantaneous signals or synchronizing the clocks with respect to another inertial frame for which there is no justification. We therefore now demand that the clocks be synchronized by slowly moving them, the meaning of which will now be made mathematically precise.

Consider the expression (1.17) for the proper time: We imagine a clock to undergo a displacement ΔX with $\Delta Y = \Delta Z = 0$, and in a time ΔT, then

$$\Delta s = \left(\Delta T^2 - 2v\,\Delta T\,\Delta X - (1 - v^2)\,\Delta X^2\right)^{\frac{1}{2}} \tag{1.23}$$

or expanding and keeping only first order displacements,

$$\Delta s = \Delta T - v\,\Delta X\,. \tag{1.24}$$

In other words the proper time of the clock changes by an amount, $-v\,\Delta X$, in addition to change in co-ordinate time. If we therefore introduce a new co-ordinate

$$t_L = T - vX\,, \tag{1.25}$$

360

this co-ordinate will incorporate the change in proper time and hence for very slow displacements $\Delta X \ll \Delta T$ we have

$$\mathrm{d}s = \mathrm{d}t_L \,. \tag{1.26}$$

If we now substitute for T, the co-ordinate t_L according to the transformation (1.25) in (1.17), and introduce $x_L = X$, $y_L = Y$, $z_L = Z$, we have

$$\mathrm{d}s^2 = \mathrm{d}t_L^2 - \mathrm{d}x_L^2 - \mathrm{d}y_L^2 - \mathrm{d}z_L^2 \,. \tag{1.27}$$

Which is once again the special relativity line element. Moreover, we note from (1.25) that

$$\begin{cases} t_L = T - vX \\ \quad = \gamma^{-1} t - v\gamma(x - vt) \\ \quad = \gamma(t - vx) \end{cases} \tag{1.28}$$

so that t_L actually does correspond to the time co-ordinate introduced in special relativity. We see, however, by this treatment that the possibility of introducing such a local time (as it was so designated by LORENTZ) depends on the possibility of making a co-ordinate transformation such as (1.25). Sometimes this is not possible as in the rotating co-ordinate frame which we shall discuss next. That is to say, we shall find that when we synchronize two clocks and slowly move one away, for example around the edge of a rotating disc, the clock upon returning to its original location will not read the same time as the clock which remained fixed, and moreover, the disagreement will depend upon the path of displacement. This is clearly not the case for (1.25) since

$$\oint \mathrm{d}t_L = \oint \mathrm{d}T \qquad \text{since} \qquad \oint v\,\mathrm{d}X = 0 \,. \tag{1.29}$$

As a consequence of not being able to satisfy a relation such as (1.29) it will not be possible to bring the line element in the rotating frame into the diagonal form of special relativity, without eliminating the rotation.

Finally we should note that the usual results of special relativity can be obtained from the line element (1.17) and co-ordinate transformation (1.16), as we have already shown for the problem of sending light signals out and back, provided we always take into account how measurements are performed. It is perhaps interesting to see in this manner how general relativity permits us to consider more complicated and varied line elements than are encountered

in the usual formulations of special relativity. The fundamental rule being that in a sufficiently small region of space and time we can, however, always « renormalize » the length of our rods, the interval of clocks, and adjust their synchronization so that the line element of special relativity holds—although in general we cannot do this for finite regions of the space-time manifold.

CHAPTER II.

ROTATING CO-ORDINATE SYSTEMS

In the preceding section we considered a transformation which did not leave the line element invariant but introduced off-diagonal terms, in particular $-2v\,\mathrm{d}X\,\mathrm{d}T$. In a rotating co-ordinate system we shall also see such terms appear, which cannot be eliminated as in the case of an inertial frame.

Apart from physical applications, rotating co-ordinate systems play an important role in general relativity for two reasons: first, in connection with the problem of the relativity of motion, that is, « does the phenomena in a rotating frame imply that rotation is absolute? » as maintained by NEWTON (the famous ice bucket experiment), « or relative? » as maintained by many philosophers contemporary with NEWTON, and most vigorously in the 19-th century by MACH—we defer a discussion of this very fundamental and difficult question to the last section. The second reason is that the spatial geometry cannot be Euclidean and this provides one of the basic arguments in general relativity for considering geometries other than Euclidean ones. The reason for this non-Euclidean structure may be seen from the following very simple argument given by EINSTEIN. Consider a rotating disc with radius R measured in the inertial frame S relative to which the disc is rotating. Now consider a small rigid rod located on the circumference of the disc. Relative to S the rod appears to be contracted because of the Lorentz contraction. Hence the number of rods one can place around the circumference will be greater than if the disc were at rest. On the other hand, a rod placed along the disc will experience no contraction since it is translating in a direction perpendicular to its orientation. Hence the circumference C' in the rotating frame will satisfy the inequality $C' > 2\pi R$ and consequently the geometry is non-Euclidean. Because of this non-Euclidean character it is *not possible* to find a spatial coordinate transformation relating the geometry of the rotating disc to the geometry in the non-rotating system. In particular if the spatial distance in S' is given by $\mathrm{d}\sigma'^2 = f(\mathrm{d}x', \mathrm{d}y', \mathrm{d}z')$ while that in the non-rotating frame is $\mathrm{d}\sigma^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$ then there does *not* exist a co-ordinate transformation

362

of the form

$$x' = f_1(x, y, z)\,, \qquad y' = f_2(x, y, z)\,, \qquad z' = f_3(x, y, z)$$

which enables us to transform the line element $d\sigma^2$ into $d\sigma'^2$.

Nevertheless, we can arrive at $d\sigma'$ by a co-ordinate transformation on ds in the following way. First of all let us write the line element in cylindrical co-ordinates, the axis of rotation being along z; then we have with

$$(2.1) \qquad \begin{cases} \varrho = (x^2 + y^2)^{\frac{1}{2}} & t = t\,, \\ \varphi = \operatorname{tg}^{-1}\dfrac{x}{y}\,, & z = z\,, \end{cases}$$

$$(2.2) \qquad ds^2 = dt^2 - d\varrho^2 - \varrho^2\, d\varphi^2 - dz^2\,.$$

To transform to the rotating co-ordinate system we employ the following transformation

$$(2.3) \qquad \begin{cases} t' = t & \varrho' = \varrho \\ \varphi' = \varphi - \omega t & z' = z\,. \end{cases}$$

The transformation in some respects resembles the Galilean transformation, and implies that the clocks in the rotating frame have been synchronized absolutely with those in the non-rotating frame. The reason for this type of co-ordinate transformation will become clear below. Then, under the above transformation, the line element takes the form

$$(2.4) \qquad ds^2 = (1 - \varrho'^2\omega^2)\, dt'^2 - 2\varrho'^2\omega\, d\varphi'\, dt' - d\varrho'^2 - \varrho'^2\, d\varphi'^2 - dz'^2\,,$$

from which we see that the proper time of a clock at rest in the rotating frame is given by

$$(2.5) \qquad ds = (1 - \varrho'^2\omega^2)^{\frac{1}{2}}\, dt\,,$$

as we would expect from special relativity. That is to say, a clock with radius arm ϱ' is travelling with speed $\varrho'\omega$, and (2.5) expresses the apparent slowing down of the clock as seen by an observer in the non-rotating frame. However, we *cannot* as in the case of uniform motion introduce a time given by

$$(2.6) \qquad T = \int (1 - \omega^2\varrho'^2)^{\frac{1}{2}}\, dt'\,,$$

since the right-hand side *cannot* be integrated, ϱ' and t' being independent

co-ordinates, alternatively stated, $\mathrm{d}T = (1-\omega^2\varrho'^2)^{\frac{1}{2}}\,\mathrm{d}t'$ does *not* define a total differential. To be sure we could define a time given by

$$T = (1-\omega^2\varrho'^2)^{\frac{1}{2}}t' \,, \tag{2.7}$$

but then $\mathrm{d}T \neq (1-\omega^2\varrho'^2)^{\frac{1}{2}}\,\mathrm{d}t'$ and the line element becomes very complicated. We therefore stay with (2.3) but realizing that t' is *not* the time read by a clock at rest in the rotating frame, and rather we must use (2.5). Note that according to (2.5) all the clocks at a given distance ϱ' from the axis of rotation have the same proper time interval as we would expect on physical grounds. Also, a disc cannot rotate with speeds $\varrho'\omega > 1$ since the proper time becomes imaginary, indeed even the case $\varrho'\omega = 1$ is forbidden since a mass particle located at this distance would, from special relativity, possess infinite energy.

On the other hand, this limitation on rotational velocities would seem to conflict with simple astronomical data. Thus, let our co-ordinate system be fixed with the earth, so that the stars are now to be regarded as in rotation about the earth, and consider the star α *Centauri A*: its distance is of the order 4.3 light years or about $4\cdot10^{18}$ cm, the earth's angular velocity ω is $7\cdot10^{-5}$ rad/s, and hence the apparent velocity of the star is $10^4\cdot c$. How is this possible?

The solution lies in observing that a system of co-ordinates attached to a rotating frame does not constitute an inertial system and special relativity makes no statement about how fast a body may travel relative to a non-inertial frame, provided one cannot use these bodies to transmit signals with speeds greater than that of light. (In particular, one cannot use observations of the fixed stars to send telegrams with speeds exceeding that of light.) On the other hand the « fixed » stars themselves apart from their random motions do constitute the fundamental inertial frame, with respect to which we understand a body such as the earth to be in rotation. (Strictly speaking the effect of nearby masses such as the moon perturbs the local inertial frame according to general relativity. The effect, however, is fortunately very small.)

We should therefore expect that seen in the rotating frame the proper time of the « fixed stars » is unchanged. To arrive at this we note that since the earth was taken as rotating with angular velocity ω relative to the stars, they consequently rotate with angular velocity $-\omega$ relative to the earth, hence for the stars, substituting

$$\mathrm{d}\varphi' = -\,\omega\,\mathrm{d}t' \tag{2.8}$$

(so that, note, they have a speed $|\varrho'(\mathrm{d}\varphi'/\mathrm{d}t')| = \varrho\omega \gg 1$) in (2.4) we find along any trajectory $\mathrm{d}\varrho = 0$, $\mathrm{d}z = 0$

$$\mathrm{d}s^2 = \mathrm{d}t'^2 \tag{2.9}$$

and the proper time is unchanged as it should be.

364

Let us now consider the propagation of light signals in the rotating frame. Setting $\mathrm{d}s^2 = 0$, we find

$$(2.10)\qquad (1-\omega^2\varrho'^2)^{\frac{1}{2}}\,\mathrm{d}t' = \frac{\varrho'^2\omega\,\mathrm{d}\varphi'}{(1-\omega^2\varrho'^2)^{\frac{1}{2}}} + \left[\mathrm{d}\varrho'^2 + \frac{\varrho'^2\,\mathrm{d}\varphi'^2}{(1-\omega^2\varrho'^2)} + \mathrm{d}z'^2\right]^{\frac{1}{2}}.$$

The expression under the radical is the proper distance between two points

$$(2.11)\qquad \mathrm{d}\sigma'^2 = \left(\mathrm{d}\varrho'^2 + \frac{\varrho'^2\,\mathrm{d}\varphi'^2}{1-\omega^2\varrho'^2} + \mathrm{d}z'^2\right)^{\frac{1}{2}}.$$

It is clearly non-Euclidean. In particular if we compute the circumference of a path at a distance R we have

$$(2.12)\qquad \int\limits_{\mathrm{d}\varrho'=\mathrm{d}z'=0} \mathrm{d}\sigma' = \frac{2\pi R}{(1-\omega^2R^2)^{\frac{1}{2}}},$$

and we see, as followed from Einstein's simple argument, the circumference C' satisfies

$$(2.13)\qquad \frac{C'}{2\pi R} = \frac{1}{(1-\omega^2R^2)^{\frac{1}{2}}} > 1\,.$$

Note that the radius is simply given by

$$(2.14)\qquad \int\limits_{\mathrm{d}\varphi'=\mathrm{d}z'=0} \mathrm{d}\sigma' = \int \mathrm{d}\varrho' = R\,,$$

in conformity with the fact that a rod oriented along the radius does not experience the Lorentz contraction. Once again we would like to stress the fact that it is *not possible* to find a transformation of the form

$$P = f_1(\varrho', \varphi', z')\,, \qquad \Phi = f_2(\varrho', \varphi', z')\,, \qquad Z = f_3(\varrho', \varphi', z')\,,$$

which eliminates the non-Euclidean structure of the element of distance $\mathrm{d}\sigma'$. Although of course we can eliminate it by employing all four co-ordinates. However, one could easily imagine a situation in which even a co-ordinate transformation between the four co-ordinates would not suffice to bring the line-element into Euclidean form. When this is the case we have a « permanent » gravitational field, such as that of the earth, the sun, etc.; in these circumstances it is not possible to eliminate the « forces » of the gravitational field everywhere by a co-ordinate transformation, although we can always make space flat « locally » as we shall show later.

Let us now consider the following experimental situation. We are on the circumference of a rotating disc and by means of mirrors we manage to send light around the circumference in opposite directions. What is the difference in time taken by the two light paths? Employing (2.10) we see that around the light path in the direction of increasing φ' we have:

$$(1-\omega^2R^2)^{\frac{1}{2}}\Delta t' = \frac{R^2\omega\cdot 2\pi}{(1-\omega^2R^2)^{\frac{1}{2}}} + \frac{R\cdot 2\pi}{(1-\omega^2R^2)^{\frac{1}{2}}}, \tag{2.15}$$

and around the path in the opposite direction

$$(1-\omega^2R^2)^{\frac{1}{2}}\Delta t' = -\frac{R^2\omega\cdot 2\pi}{(1-\omega^2R^2)^{\frac{1}{2}}} + \frac{R\cdot 2\pi}{(1-\omega^2R^2)^{\frac{1}{2}}}. \tag{2.16}$$

Hence subtracting and denoting $(1-\omega^2R^2)^{\frac{1}{2}}\Delta t'$ by $\Delta S'$ the proper time on the circumference, we have

$$\underset{\circlearrowleft}{\Delta S'} - \underset{\circlearrowright}{\Delta S'} \doteq \frac{4\pi R^2\omega}{(1-\omega^2R^2)^{\frac{1}{2}}}. \tag{2.17}$$

Thus, apart from 2-nd order effects, the time difference is given by, restoring units,

$$4\times(\text{area of path})\times\omega/c^2\,. \tag{2.18}$$

This effect has actually been measured both for a rotating platforn, where it was possible to control ω, by SAGNAC in 1914, and for the earth by MICHELSON and GALE in Chicago in 1925. In the case of the Michelson-Gale experiment light was sent around a rectangle of approximately a kilometer on a side and the difference in time measured by an interferometer technique. The experiment may be looked upon as the optical analogue of the Foucault pendulum, whereby one determines the earth is in rotation by observing the precession of a pendulum without direct reference to the fixed stars.

Let us now return to the problem of the slowly-moved clocks. Instead of sending light signals around the circumference, let us consider two clocks: one, fixed at the point A on the rotating disc, and the other synchronized with the clock at A and now slowly displaced around the circumference of radius R. Then the proper time of the slowly displaced clock is given by:

$$\Delta S = (1-\omega^2R^2)^{\frac{1}{2}}\Delta t' - \frac{R^2\omega\,\Delta\varphi'}{(1-\omega^2R^2)^{\frac{1}{2}}}. \tag{2.19}$$

And hence after displacing by $\Delta\varphi' = 2\pi$ the clock upon returning to A differs

366

from the reading of the clock at A by (restoring units).

$$\Delta S_{\circlearrowleft} - \Delta S_A = -\frac{2\pi R^2 \omega c^{-2}}{(1-\omega^2 R^2)^{\frac{1}{2}}} . \tag{2.20}$$

So that the Michelson-Gale experiment could equally as well have been performed using slowly-moved clocks. For example, if we take a path around the equator $R = 6.4 \cdot 10^3$ km, neglecting the 2-nd order term, we have $\Delta S_{\circlearrowleft} - \Delta S_A \approx 10^{-7}$ s, which could easily be measured, although because of the information we now have from other experiments, it would only serve as a further check of the basic ideas.

Chapter III.

THE PRINCIPLE OF EQUIVALENCE

Perhaps one of the most fundamental questions one can ask concerning the general theory of relativity is, « Why does the theory concern itself so much with gravitation? » We have given a partial answer to this in the Preface, but we now wish to develop this point more fully.

As is well known in Newtonian mechanics and gravitational theory, the acceleration imparted to a particle by a gravitational field is independent of the mass of the particle. We have, for example, in a uniform field in the negative z-direction, for a particle of mass m freely falling in it,

$$m\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -mg , \tag{3.1}$$

$$m\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = m\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = 0 . \tag{3.2}$$

Or we may simply write for (3.1)

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -g . \tag{3.3}$$

The curious thing about this equation is that the mass does not enter in it because of the equality of gravitational and inertial mass. This is quite dif-

ferent than the case of the Lorentz force: for a particle of charge e and mass m in a uniform electrostatic field $\mathscr{E}_z$, the equations are (neglecting relativistic corrections)

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = m\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = 0\,, \qquad m\frac{\mathrm{d}^2z}{\mathrm{d}t^2} = e\mathscr{E}_z\,. \tag{3.4}$$

We see the acceleration imparted to the charge depends on the ratio e/m. As a consequence if we have a collection of charged particles of various e/m ratios, even neglecting their mutual interactions, they will show a tendency to separate in the presence of the uniform field $\mathscr{E}_z$. This forms the basis of a simple e/m spectrometer. On the other hand, if we drop from a given height at the same time, a number of particles having different masses m_1, m_2, m_3, ... etc., they will strike the earth at the same time, showing no tendency to separate. As a consequence if we adopt a system of reference accelerated downward with an acceleration g, the particles appear at rest in this frame, and do not appear to be acted on by any forces of the gravitational field.

Consider now a rocket ship in free space not under power; it obviously constitutes an inertial frame. A particle of mass m in the rocket ship therefore satisfies the equations

$$m\frac{\mathrm{d}^2z}{\mathrm{d}t^2} = 0\,, \quad \text{etc.}\,. \tag{3.5}$$

Let the rocket be turned on, so that the rocket ship undergoes a uniform acceleration in the z direction, corresponding to the co-ordinate transformation

$$z' = z - \tfrac{1}{2}at^2 \tag{3.6}$$

and hence

$$\frac{\mathrm{d}^2z'}{\mathrm{d}t^2} = \frac{\mathrm{d}^2z}{\mathrm{d}t^2} - a\,. \tag{3.7}$$

Substituting (3.7) in (3.5) we have

$$m\frac{\mathrm{d}^2z}{\mathrm{d}t^2} = -\,ma \qquad \text{or} \qquad \frac{\mathrm{d}^2z}{\mathrm{d}t^2} = -\,a\,. \tag{3.8}$$

Thus the acceleration of the particle in the rocket ship would be independent of its mass as in the gravitational field. Moreover, if we imagine the particle attached to a spring scale it would appear to have acquired a weight $m\,a$ just as though the rocket ship were at rest on the surface of the earth and the particle was being « pulled down » by the gravitational attraction. (Indeed it is

368

this property of inertia which has led to the suggestion of introducing an effective gravity into artificial satellites, *e.g.* by rotating them.)

We see therefore that because of the principle of equivalence there is a kind of relativity of acceleration. An observer whose accelerometer says he is undergoing a uniform acceleration cannot distinguish whether he is

a) in free space accelerating relative to the « fixed stars »,

b) at rest on the surface of a gravitating body.

And a similarly when his accelerometer reads zero, he cannot distinguish whether he is

a) uniformly translating in outer space relative to the stars,

b) freely falling in the earth's gravitational field and hence undergoing an accelerated motion relative to the stars.

These important results follow from Newton's laws of motion and the equivalence of inertial mass and gravitational mass.

However, if we now consider special relativistic effects, since the energy of a body increases when it is in motion relative to a given frame, and hence its inertial mass (from the equivalence of inertial mass and energy via the famous relation $m = E/c^2$), it would follow that unless the gravitational mass also increased, we would no longer have the principle of equivalence holding exactly. (It is an interesting feature of Einstein's approach that he wished to keep this very fundamental property of the Newtonian theory, just as in developing special relativity he kept the relativity of uniform motion.)

This distinction would lead to an inconsistency with the conservation of energy which we may arrive at by the following argument.

From special relativity, if a system's energy changes by ΔE its inertial mass also changes by $\Delta m = \Delta E/c^2$. Hence if we imagine a nucleus at rest in the gravitational field with mass M_0, and consequently gravitational mass M_0, to undergo a transition in which it emits a photon of energy E to a final state $M_F = M_0 - E$ (neglecting recoil, diminution of recoil is of importance in connection with the Mössbauer effect and experimental verification of the principle of equivalence), it will have after the transition a gravitational mass M_F as may be verified experimentally. We therefore expect that the gravitational mass of a body will depend on its energy content. Now since the emitted photon has an energy $E = h\nu$, it should therefore have a gravitational mass $m = h\nu/c^2$, and therefore upon « falling » in a uniform gravitational field of strength g, acquire an additional energy $(h\nu/c^2)gL$, where L is the distance it « falls ». Its final energy will therefore be

$$h\nu + \frac{h\nu}{c^2}\, gL = h\nu\left(1 + \frac{gL}{c^2}\right). \tag{3.9}$$

If this were not the case we would arrive at the following contradiction with the law of conservation of energy. Let us imagine the above nucleus in the initial state M_0 to fall a distance L before decaying, then it will acquire an additional energy $M_0 gLc^2$ and its total energy before decaying will be

$$E_1 = M_0c^2 + M_0gLc^2 \,. \tag{3.10}$$

Now consider the energy obtained if the nucleus first decays and then falls, the radiation will also be assumed to be received at distance L below its point of emission. If the photon did not acquire an additional energy given by (3.9), the total energy received would be

$$E_2 = M_Fc^2 + M_FgLc^2 + h\nu \,. \tag{3.11}$$

Subtracting (3.11) from (3.10) and using $h\nu = (M_0 - M_F)c^2$ we have

$$E_1 - E_2 = (M_0 - M_F)gLc^2 \,. \tag{3.12}$$

On the other hand, if we take the photon's energy to be (3.9), so that instead of (3.11) we have

$$E_2' = M_Fc^2 + M_FgLc^2 + h\nu\left(1 + \frac{gL}{c^2}\right), \tag{3.13}$$

and we have $E_1 - E_2' = 0$, so that energy is conserved.

We shall now show how we may obtain (3.9) from an elementary application of the Doppler effect and the possibility of replacing a system of co-ordinates at rest in a uniform gravitational field by one in uniform acceleration. Consider two observers A and B at rest in a uniform gravitational field of magnitude g with B located vertically above A at distance L. Now let a photon of frequency ν be emitted at B and absorbed at A. What is the frequency of the photon measured at A and what is its energy?

According to our basic principle discussed previously the system is equivalent to one in which there is no gravitational field present, but the frame in which A and B are at rest is undergoing a uniform acceleration in the positive z-direction. Now consider a photon emitted from B with frequency ν_B, then in travelling from B to A, the observer A has in the meantime experienced an acceleration lasting a time $t_{BA} = L/c$ (approximately), and has acquired a velocity $V = gL/c$ and hence by the Doppler effect the frequency of the photon

370

seen by the observer at A will not be ν_B, but

$$\nu_A = \nu_B \left(1 + \frac{v}{c}\right) = \nu_B \left(1 + \frac{gL}{c^2}\right), \tag{3.14}$$

and hence the energy will be $E_A = h\nu_A = E_B(1 + (gL/c^2))$ as in (3.9). Thus the photon will have acquired energy as a consequence of its fall in the gravitational field, and similarly on rising it will lose energy.

This result will enable us to make a crude estimate of the deflection of light in a gravitational field; however, before doing this let us call attention to an important consequence of (3.14).

If we imagine an oscillator sending out a wave train from B to A, then in time Δt_B the number of waves which has left B must be $\nu_B \Delta t_B$; this wave train will appear at A during time Δt_A, and the number of waves will be $\nu_A \Delta t_A$. Since the total number of waves is clearly the same, we must have

$$\nu_A \Delta t_A = \nu_B \Delta t_B \tag{3.15}$$

TABLE I. – *Red-shift of 0-type stars in stars clusters.*
« These stars may have masses of the order of 100 solar masses and radii about 4 ÷ 5 times that of the sun. The gravitational red-shift would thus be about 20 times greater than on the sun or near 10 km/s which is about the average value observed » (from R. J. TRUMPLER).

Cluster	Star	M	Sp. T.	Red-Shift (km/s)
J.C. 1805	2	-4.5	07	$+12.4 \pm 2.2$ S.B.
	3	-4.2	07	$+\ 2.8 \pm 2.0$
J.C. 1848	1	-6.0	07	$+10.4 \pm 4.8$ S.B.
	1*a*	-4.4	08	$+\ 4.6 \pm 5.4$
	2	-4.9	09	$+\ 6.2 \pm 3.5$
NGC. 2244	9	-5.0	07	$+\ 6.8 \pm 1.3$
	15	-4.6	06	$+13.6 \pm 1.7$
	8	-4.3	08	$+\ 6.4 \pm 1.6$
NGC. 2264	1	-4.6	07	$+\ 9.8 \pm 1.2$
NGC. 2353	1	-4.6	09	$+16.1 \pm 1.6$ S.B.
NGC. 6231	50	-4.6	08	$+16.4 \pm 2.6$
NGC. 6604	1	-6.0	08	$+13.6 \pm 4.1$ Var. Vel.
NGC. 6611	1	-5.2	06	$+\ 9.0 \pm 2.1$
	2	-4.8	08	$+\ 4.1 \pm 3.7$ Var. Vel.
	3	-4.7	08	$+\ 9.9 \pm 2.4$
NGC. 6823	1	-4.3	08	$+11.6 \pm 3.4$
	2	-4.0	09	$+\ 7.7 \pm 3.6$
NGC. 6871	5	-5.4	B0	$+15.6 \pm 1.6$
18 stars in 10 clusters				$+\ 9.8$ km/s

371

and hence using (3.9), it follows

$$\Delta t_A \left(1 + \frac{gL}{c^2}\right) = \Delta t_B . \tag{3.16}$$

From this we can conclude that if two clocks are at rest in a gravitational field, they will run at different rates depending on their location. In particular a clock at the surface of the earth will run more slowly than a clock located on a mountain.

The expression (3.14) has been verified astronomically in the case of the sun and a variety of stars, although there is some experimental uncertainty in the result, the accompanying Table I taken from the talk of R. J. TRUMPLER (*Jubilee of Relativity Theory*, in *Helv. Phys. Acta*, Supplementum IV, Basel, 1956) would seem to leave no doubt that the effect is definitely present and of the right order of magnitude.

In addition to the above measurements, recently using the Mössbauer effect, the shift has been measured terrestrially most accurately by R. V. POUND and G. A. REBKA jr., *Phys. Rev. Lett.*, **4**, 337 (1960). Since the height L of their apparatus was ≈ 25 m the frequency shift was of order 10^{-15} in comparison with the shift of order $3\cdot 10^{-5}$ in the above table. Thus the shift has been checked experimentally over a range of 10^{10} in $\Delta\nu/\nu$.

The above formulas are based on a uniform gravitational field; however, a natural generalization is to replace gL by ΔU, the difference in gravitational potential between B and A. Hence, if we choose the arbitrary constant in the potential such that it is zero at infinity, it follows from (3.16) that the time interval of a clock in a gravitational field in terms of the time interval Δt of a clock at infinity is

$$\Delta t_A = \left(1 + \frac{U(A)}{c^2}\right) \Delta t . \tag{3.17}$$

That is we have $\Delta t_A \approx (1 - (gL/c^2))\, \Delta t_B$ (our calculation is not valid to higher order terms). Now inserting the expression for the Newtonian gravitational potential, $U_A = - GM/r$, where M is the mass of the body,we have

$$\Delta t_A = \left(1 - \frac{GM}{rc^2}\right) \Delta t . \tag{3.18}$$

An expression we shall obtain later in a more rigorous way.

Now it follows from the basic principle we employed earlier, that since

Δt_A represents the time read by a clock at rest at A, this must be the proper time, that is for a clock at rest

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2\,\mathrm{d}t^2\,, \tag{3.19}$$

apart from higher order terms. Therefore to a *first approximation* we would expect

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2\,, \tag{3.20}$$

and we see that the velocity of light, obtained from $\mathrm{d}s=0$, is no longer « c » over finite regions. Because of this dependence of the velocity on the location, a ray of light passing near a body of mass M will be deflected. From a Newtonian standpoint, this can also be understood in the following way. We associate with a photon of frequency $h\nu$, a gravitational mass $h\nu/c^2$ as above, then on passing by the body of mass M it is attracted by the gravitational field, just as is a Newtonian particle. However, as we indicated earlier this picture is misleading, since the geodesic paths followed by a body, *i.e.* $\delta\int \mathrm{d}s = 0$ do not represent paths for which the body « feels » a force. The error lies in extrapolating from the Euclidean geodesics of flat space, or regions where the gravitational field is negligibly small, to regions where the geodesics have become significantly non-Euclidean, and then regarding these departure from the Euclidean geodesics as being due to « gravitational forces ».

To calculate the deflection of light we should therefore solve the equations

$$\begin{cases} \delta\displaystyle\int \mathrm{d}s = 0\,, \\ \mathrm{d}s = 0\,. \end{cases} \tag{3.21}$$

However, the exact method of solving such equations belongs to a later chapter; moreover, since our expression for $\mathrm{d}s$ is only approximate, such a solution is not necessary. Rather let us proceed in the following way. We have seen that a photon in a gravitational field behaves very much as though it were a Newtonian particle of mass $m = h\nu/c^2$ travelling with speed c, *i.e.*, a momentum $h\nu/c$. Let us therefore calculate the deflection experienced by a Newtonian particle travelling with speed c in a Newtonian gravitational field. It may be shown this gives the same answer as one obtains from (3.21), although strictly speaking of course, such a picture is untenable from the standpoint of Huyghen's

principle. Indeed, the effect of the gravitational field is to increase the index of refraction in the region of space so that light gets deflected towards the gravitational field, and therefore the speed of light is *less* near the gravitational field than at great distances. On the other hand, a Newtonian particle as it approaches the sun, or gravitating bodies, *increases* its speed. It is for this reason that when EINSTEIN calculated the effect he was very careful to use Huyghen's principle and *not* the method we shall present below. The student should read for himself this very elegant and succinct derivation of the effect by EINSTEIN.

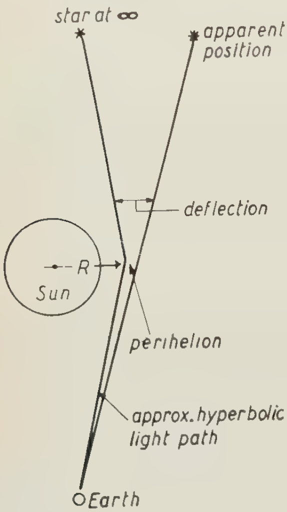


Fig. 1.

The physical situation is given in the above picture. We imagine now a Newtonian particle of mass m moving in a plane with polar co-ordinates r, φ. The energy integral is

$$\frac{1}{2}\dot{r}^2 + \frac{1}{2}r^2\dot{\varphi}^2 - \frac{GM}{r} = \frac{E}{m}, \tag{3.22}$$

where $\dot{r}$, $\dot{\varphi}$ denote radial and angular velocities, M is the mass of the deflecting body, E is the energy, m the mass of the particle. The angular momentum integral is (Kepler's law)

$$r^2\dot{\varphi} = \frac{A}{m}, \tag{3.23}$$

where A is the angular momentum. To estimate A, we note that when the particle is at the perihelion $r\dot{\varphi} \approx c$. We therefore set $A = mRc$ and hence

$$r^2\dot{\varphi} = Rc\,. \tag{3.24}$$

Hence, we may eliminate time derivatives above using

$$\frac{\mathrm{d}}{\mathrm{d}t} = \dot{\varphi}\frac{\mathrm{d}}{\mathrm{d}\varphi} = \frac{Rc}{r^2}\frac{\mathrm{d}}{\mathrm{d}\varphi},$$

to write (3.22) as

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\frac{R^2c^2}{r^4} + \frac{1}{2}\frac{R^2c^2}{r^2} - \frac{GM}{r} = \frac{E}{m}, \tag{3.25}$$

and now introducing $u = r^{-1}$, the above becomes

$$\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 R^2c^2 + \frac{1}{2}R^2c^2u^2 - GMu = \frac{E}{m}, \tag{3.26}$$

which upon differentiation yields

$$\frac{\mathrm{d}^2u}{\mathrm{d}\varphi^2}+u=\frac{GM}{c^2R^2}\,, \tag{3.27}$$

Newton's equation for planetary motion in polar co-ordinates for $A=mRc$. Now the solution to this equation consists of the solution to the homogeneous equation $\mathrm{d}^2u/\mathrm{d}\varphi^2+u=0$, or the straight line $r\cos\varphi=R$

$$u=\frac{1}{R}\cos\varphi\,, \tag{3.28}$$

plus the inhomogeneous part $u=GM/c^2R^2$ corresponding to pure circular motion $r=\text{constant}$. The total solution is (very nearly)

$$u=R^{-1}\cos\varphi+GM/c^2R^2\,, \tag{3.29}$$

since $R^{-1}\gg GM/c^2R^2$ the above is the equation of a hyperbola (for a planet the second term is much larger than the first corresponding to an ellipse). Now what we desire is the angular deviation of the two asymptotes to the hyperbola. Therefore since these asymptotes correspond to $u=0$, we have

$$\begin{cases} R^{-1}\cos\varphi_1+GM/c^2R^2=0\,,\\ R^{-1}\cos\varphi_2+GM/c^2R^2=0\,. \end{cases} \tag{3.30}$$

Now since φ_1 and φ_2 differ only by a small amount from $\pm\pi/2$ and are clearly equal in magnitude, we have, using $\cos(\Delta+(\pi/2))=-\sin\Delta\approx-\Delta$

$$\Delta\approx GM/c^2R \tag{3.31}$$

and the angular deflection from the lower asymptote, projected backwards, is

$$\alpha=2\Delta\approx 2GM/c^2R\,. \tag{3.32}$$

This result is too small by a factor of two, and to get the correct answer one must resort to the full machinery of the general theory. Although SCHIFF (*American Journal of Physics*, **28**, 340 (1960)) has given some heuristic arguments recently using the principle of equivalence to a higher order than we have here to obtain the correct deflection without using the complete theory (*).

(*) See also N. L. BALAZS: *Zeits. Phys.*, **154**, 264 (1959), and the work of W. LENZ (1944) described in A. SOMMERFIELD: *Lectures on Theoretical Physics*, vol. III (New York, 1952).

875

However, it should be understood that once we adopt a line element different from that of special relativity (*i.e.*, one that cannot be obtained by a co-ordinate transformation from $ds^2 - dt^2 - dx^2 - dy^2 - dz^2$) we are no longer in the framework of Minkowskian geometry or of special relativity and one does not have an explanation of the phenomena unless one has a well-defined mathematical scheme for arriving at such a non-Euclidean line element. Thus it should be remembered that EINSTEIN used such elementary arguments as we have given above *before* he had constructed the general theory and as a guide to what the theory should contain, so that such arguments should be looked upon as bearing about the same relation to the general theory as Bohr's quantum rules bear to quantum mechanics. We therefore proceed to the construction of the full theory, which in addition to giving us the change in clock rate, deflection of light, will give us the precession of the perihelion of Mercury.

Finally in connection with the experimental side of the principle of equivalence, in addition to the above references, we have the famous work of the Hungarian physicist R. VON EÖTVÖS (*Ann. of Phys.*, **59**, 354 (1896)) who demonstrated experimentally the equivalence of inertial mass and gravitational mass to about one part in 10^8. This was a considerable improvement over the work of BESSEL (around 1830), based on considering the period of pendulum. For a simple pendulum one has from Newton's laws $T = 2\pi\sqrt{l/g}$, where l is the length of the pendulum, but if inertial mass and gravitational mass were not equal one would have $T = 2\pi\sqrt{m_I l/m_g g}$, where m_I, m_g are the inertial and gravitational masses respectively. Clearly as we change m_g unless m_I changes proportionately, the period will also change. By this method Bessel verified the principle of equivalence to a few parts in 10^5. Recently R. H. DICKE of Princeton has been engaged in a program to improve on the results of EÖTVÖS and to see whether or not the principle is correct to a higher order of accuracy.

CHAPTER IV.

TENSOR CALCULUS AND RIEMANNIAN GEOMETRY

4·1. – The requirement of general covariance.

In the preceding sections we saw that the quadratic differential form or line element ds^2 provided us with a convenient tool for studying the propagation of light in different co-ordinate systems. In performing these co-ordi-

376

nate transformations and in writing down the line element no special notation was used, in order to avoid confusing mathematical notation with the physical ideas involved. However, although it is possible to continue in the preceding fashion, we enormously simplify the presentation by introducing covariant notation and the calculus of tensors. We shall not attempt to give a rigorous mathematical presentation of this discipline, but rather develop it side by side with the physical ideas.

As you know the fundamental postulate of special relativity is:

> *The laws of nature take the same form in each of two uniformly moving frames.*

As a consequence of this postulate one is restricted in special relativity to considering co-ordinate transformations which leave ds^2 form-invariant. These transformations constitute the group of Lorentz transformations, and in special relativity one is restricted to this group of transformations. However in general we are not merely confronted with uniformly moving frames, but frames in arbitrary states of motion, for which there is consequently no Lorentz transformation connecting the two frames. How are we to formulate the laws of nature in such frames? Moreover frequently we are interested in using other systems of spatial co-ordinates besides those which leave ds^2 form-invariant. For example in Chapter II, we introduced cylindrical co-ordinates for which the line element assumed the form

$$ds^2 = dt_2 - d\varrho^2 - \varrho^2 d\varphi^2 - dz^2 \,. \tag{4.1}$$

Such a transformation, by definition, is not a Lorentz transformation, nevertheless we can certainly describe how light propagates in such a co-ordinate system. Indeed Maxwell's equations written in vector form

$$\begin{cases} \nabla\cdot\boldsymbol{D} = \varrho & \nabla\times\boldsymbol{\mathscr{H}} = \boldsymbol{J} + \partial_t\boldsymbol{D} \\ \nabla\cdot\boldsymbol{B} = 0 & \nabla\times\boldsymbol{\mathscr{E}} + \partial_t\boldsymbol{B} = 0 \end{cases} \tag{4.2}$$

make no reference whatsoever to the system of spatial co-ordinates employed. They hold in arbitrary systems of spatial co-ordinates.

On the other hand we have learned from special relativity that time plays the role of an additional co-ordinate, so that we are no longer dealing with a three dimensional manifold and an absolute time as in Newtonian mechanics, but a four dimensional manifold, the Euclidean element of spatial distance $d\sigma$ and the element of time being combined to form a four dimensional element of distance

$$ds^2 = dt^2 - d\sigma^2 \,. \tag{4.3}$$

However because of the difference in sign of the way dt^2 and $d\sigma^2$ occur in ds^2, it is clear that they do not form a 4-dimensional space in the Euclidean sense which would correspond to a line element of the form

$$ds^2 = dt^2 + d\sigma^2 \, . \tag{4.4}$$

This distinction is very important since in the latter case, the propagation of light would be described by $dt^2 + d\sigma^2 = 0$ and hence $dt = 0$, $d\sigma = 0$ in other words light would not propagate (*). On the other hand setting (4.3) equal to zero, we have $\Delta t = \Delta\sigma$ correspond to light propagating through a distance $\Delta\sigma$ in a time Δt. Apart from this distinction, however, we may look upon time as simply an additional co-ordinate that enters into the transformation along with the three spatial co-ordinates.

We are therefore led by the above arguments to the conclusion that the laws of nature should be stated in a form that hold under arbitrary co-ordinate trasformations, treating time as a co-ordinate along with the spatial co-ordinate. (Postulate of general covariance.) The mathematical tool for accomplishing this will be the covariant tensor calculus invented by RICCI to handle precisely this kind of problem.

4·2. – Covariant notation.

As we have seen the line element under a co-ordinate transformation assumed a variety of different values for the coefficients of the dt^2, $dt\,dx$, dx^2, dy^2, dz^2 etc. To describe this situation we write it in the following notation, and we sum over repeated indices designating the four co-ordinate differentials by dx^μ $(\mu = 0, 1, 2, 3)$

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu \, . \tag{4.5}$$

The $g_{\mu\nu}$ is called the metric tensor: *metric*, because it provides the relation between the co-ordinate differentials which furnish the differential element of distance (**) between events, and *tensor* because of the way in which it behaves under co-ordinate transformations. The reason for writing subscripts below on the metric tensor, and above on the co-ordinate differentials also arises

(*) We are of course excluding imaginary values for the co-ordinates. Alternatively if we use Minkowski's trick of setting $x_4 - it$ we exclude *real* values of x_4.

(**) It is of course possible to consider generalized spaces in which there is no element of distance defined, as is done by the topologists. Also, needless to say, although we have restricted ourselves to four co-ordinates the notation itself involves no such restriction.

because of the way in which they transform. As we have seen the $g_{\mu\nu}$ are in general functions of the co-ordinates and not simply constants as in special relativity.

On the other hand we have seen from our earlier discussion that in general what a clock or rod measures does not depend only on the co-ordinates but on the $g_{\mu\nu}$ as well. Indeed as we have seen the proper time of a clock at rest at a particular point will not be $\mathrm{d}x^0$, but

$$\mathrm{d}s = \sqrt{g_{00}(\mathrm{d}x^0)^2}\,. \tag{4.6}$$

It follows that we must restrict ourselves on physical grounds to co-ordinate transformations for which (*)

$$g_{00} > 0\,. \tag{4.7}$$

Similarly upon setting $\mathrm{d}s^2 = 0$, we have for the proper time to propagate through Δx^i,

$$\left\{\begin{aligned} &g_{00}^{\frac{1}{2}}\Delta x^0 = -\,g_{0i}g_{00}^{-\frac{1}{2}}\Delta x^i + \sqrt{\gamma_{ij}\Delta x^i\Delta x^j}\,, \qquad (i, j = 1, 2, 3)\\ &\gamma_{ij} = \frac{g_{0i}g_{0j}}{g_{00}} - g_{ij}\,. \end{aligned}\right. \tag{4.8}$$

The γ_{ij} represent the metric for the element of spatial distance, as can be seen if we consider a situation in which

$$\oint \frac{g_{0i}\Delta x^i}{g_{00}^{\frac{1}{2}}} = 0\,. \tag{4.9}$$

By our previous arguments a rod of co-ordinate differential Δx^1 does not have a proper length Δx^1, but a proper length

$$\sqrt{\gamma_{11}(\Delta x^1)^2}\,. \tag{4.10}$$

For example in cylindrical co-ordinates a rod located along $\Delta\theta$, doesn't have length $\Delta\theta$, but $r\,\Delta\theta$. Since on physical grounds the proper distance between two points should be positive definite and locally reducible to the Euclidean

(*) Sometimes in the literature one finds this requirement reversed *i.e.* $g_{00} < 0$; this is because the authors define their $\mathrm{d}s^2$ as the negative of our $\mathrm{d}s^2$. Also some authors have used « light-cone » co-ordinates in the discussion of gravitational waves; in these co-ordinates $g_{00} = 0$. The above restriction cannot be maintained everywhere in the gravitational field of a point mass for some co-ordinate systems. Thus the above restriction may have to be qualified with further development of the theory, or, alternatively, it will serve to greatly restrict the class of allowed solutions, and energy-momentum tensors.

distance, the quadratic form satisfies

$$\gamma_{ij}\, \Delta x^i \Delta x^j > 0 \tag{4.11}$$

unless all Δx^i. Hence the 3×3 matrix γ_{ij} must be positive definite

$$\gamma_{11} > 0 \qquad \begin{vmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{vmatrix} > 0 \qquad \begin{vmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{vmatrix} > 0\,. \tag{4.12}$$

We are therefore restricted to co-ordinate systems for which (4.7) and (4.12) hold. For example if we try to transform time into a space co-ordinate and space co-ordinate into time (4.7) and (4.12) are violated. By the above method we therefore arrive at our systems of allowed co-ordinates, it being understood that, in general, co-ordinates do not have any physical significance but only certain « proper » measures of them such as in (4.6) and (4.10).

Let us now adopt a $g_{\mu\nu}$ satisfying the above restrictions and perform a co-ordinate transformation, and see what form it takes under the transformation. We have under a co-ordinate transformation from $\mathrm{d}x^\mu$ to $\mathrm{d}x'^\nu$

$$\mathrm{d}x^\mu = \left(\frac{\mathrm{d}x^\mu}{\mathrm{d}x'^\lambda}\right)\mathrm{d}x'^\lambda, \tag{4.13}$$

and hence substituting in $\mathrm{d}s^2$

$$\mathrm{d}s^2 = g_{\mu\nu}\left(\frac{\mathrm{d}x^\mu}{\mathrm{d}x'^\lambda}\right)\left(\frac{\mathrm{d}x^\nu}{\mathrm{d}x'^\varrho}\right)\mathrm{d}x'^\lambda\,\mathrm{d}x'^\varrho\,. \tag{4.14}$$

So that the metric tensor in the primed frame is given by

$$g'_{\mu\nu} = \left(\frac{\partial x^\mu}{\partial x'^\lambda}\right)\left(\frac{\partial x^\nu}{\partial x'^\varrho}\right) g_{\mu\nu}\,. \tag{4.15}$$

Quantities which transform in this fashion are called covariant tensors of the second rank (two indices). We see also that since $g_{\mu\nu} = g_{\nu\mu}$ it is a symmetric tensor. On the other hand, the $\mathrm{d}x'^\lambda$ transforms from (4.13) as

$$\mathrm{d}x'^\lambda = \left(\frac{\partial x'^\lambda}{\partial x^\mu}\right)\mathrm{d}x^\mu\,. \tag{4.16}$$

We say it transforms as a contravariant tensor of the first rank, or contravariant *vector*.

Remark. one should carefully distinguish between the transformation prop-

erties of the co-ordinate *differentials* and the co-ordinates themselves. Thus x'^λ is not a tensor (even though our notation would imply this (*)) since

$$x'^\lambda = f^\lambda(x^1, x^2, x^3, x^4) \tag{4.17}$$

and this function is not in general the same as (4.16); only in the case of *linear*, homogeneous transformations are the two transformations the same, An example of a quantity transforming as a covariant tensor of the first rank is furnished by considering the gradient of a scalar function of the co-ordinates, say Φ. Thus

$$\frac{\partial \Phi}{\partial x'^\mu} = \frac{\partial \Phi}{\partial x^\nu}\left(\frac{\partial x^\nu}{\partial x'^\mu}\right), \tag{4.18}$$

or denoting $\partial\Phi/\partial x'^\mu$ by V'_μ we have

$$V'_\mu = \left(\frac{\partial x^\nu}{\partial x'^\mu}\right) V_\nu . \tag{4.19}$$

Any vector transforming as in (4.19) we say transforms as a covariant tensor of the first rank.

Now consider the expression $V'^\mu V'_\mu$ under a co-ordinate transformation

$$V'^\mu V'_\mu = \left(\frac{\partial x'^\mu}{\partial x^\lambda}\right)\left(\frac{\partial x^\nu}{\partial x'^\mu}\right) V^\lambda V_\nu . \tag{4.20}$$

But clearly since

$$\mathrm{d}x^\nu = \frac{\partial x^\nu}{\partial x'^\mu}\,\mathrm{d}x'^\mu = \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial x'^\mu}{\partial x^\lambda}\,\mathrm{d}x^\lambda , \tag{4.21}$$

we have

$$\frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial x'^\mu}{\partial x^\lambda} = \delta^\nu_\lambda , \qquad \left\{\begin{matrix} = 1, & \nu = \lambda \\ = 0, & \nu \neq \lambda \end{matrix}\right\}, \tag{4.22}$$

where δ^ν_λ is the Krönecker-δ symbol. It is a « mixed » tensor and is the same for all observers since

$$\delta'^\lambda_\varrho = \left(\frac{\partial x'^\lambda}{\partial x^\mu}\right)\left(\frac{\partial x^\nu}{\partial x'^\varrho}\right)\delta^\mu_\nu = \frac{\partial x'^\lambda}{\partial x^\mu}\frac{\partial x^\mu}{\partial x'^\varrho} = \delta^\lambda_\varrho . \tag{4.23}$$

(*) Some authors attempt to avoid this by writing the co-ordinates with the subscript below *i.e.* x'_λ, but then since x'_λ is not a covariant tensor this does not make for greater logical consistency than writing it above which greatly facilitates the notation.

Hence

$$V'^{\mu}V'_{\mu} = \delta^{\nu}_{\lambda}V^{\lambda}V_{\nu} = V^{\nu}V_{\nu}\,. \tag{4.24}$$

An expression such as $V^{\nu}V_{\nu}$ is described as an invariant scalar product, it is a tensor of rank zero since denoting it by S

$$S' = V'^{\mu}V'_{\mu} = V^{\nu}V_{\nu} = S\,. \tag{4.25}$$

We see from the above that tensors of higher or lower rank may be formed by multiplication of tensors (it being understood that the argument of the tensors in the product is taken at the same point), for example, we can form the « mixed » tensor

$$T^{\mu\nu\lambda\varrho\dots}_{\alpha\beta\gamma\delta\dots} = A^{\mu}B^{\nu}C^{\lambda}D^{\varrho}\dots E_{\alpha}F_{\beta}G_{\gamma}H_{\delta}\dots\,. \tag{4.26}$$

So far we have not directly related covariant and contravariant indices. We define for a contravariant vector A^{ν}, the covariant vector given by

$$A_{\mu} = g_{\mu\nu}A^{\nu}\,. \tag{4.27}$$

Then clearly $A^{\mu}A_{\mu} = g_{\mu\nu}A^{\nu}A^{\mu}$ in analogy with the line element $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}$. Similarly given a second rank contravariant tensor $T^{\mu\nu}$, we may define

$$T_{\mu\nu} = g_{\mu\lambda}g_{\nu\varrho}T^{\lambda\varrho} \tag{4.28}$$

and so on. As may be easily verified A_{μ}, $T_{\mu\nu}$ defined in this way transform as covariant tensors of the first and second ranks, and moreover $A_{\mu}A^{\mu}$, $T_{\mu\nu}T^{\mu\nu}$ are invariant tensors.

As may be easily shown from our assumptions (4.7) and (4.12) the determinant of $g_{\mu\nu}$ cannot vanish (since $\det g_{\mu\nu} = g = -g_{00}\gamma$ and g_{00}, $\gamma > 0$) (*) hence we may define an inverse to $g_{\mu\nu}$, the contravariant metric tensor $g^{\lambda\varrho}$ which satisfies

$$g^{\lambda\varrho}g_{\varrho\nu} = \delta^{\lambda}_{\nu}\,. \tag{4.29}$$

Since by definition the matrix with components $M^{\varrho\nu}$ the so-called minor to $g_{\varrho\nu}$ satisfies

$$g_{\varrho\nu}M^{\varrho\lambda} = g\delta^{\lambda}_{\nu} \tag{4.30}$$

we have $g^{\lambda\varrho} = M^{\lambda\varrho}/g$.

(*) Apart from exceptional cases, such as polar co-ordinates at the origin.

4·3. – Tensor densities.

In addition to quantities transforming as tensors we in general encounter quantities transforming as

$$T'^{\mu\dots}_{\lambda\dots} = \left|\frac{\partial x}{\partial x'}\right|^{W} \left(\frac{\partial x'^{\mu}}{\partial x^{\zeta}}\right)\left(\frac{\partial x^{\varrho}}{\partial x'^{\lambda}}\right) \dots T^{\zeta\dots}_{\varrho\dots}\,, \tag{4.31}$$

where W is a number and $|\partial x/\partial x'|$ the Jacobian. Such quantities as T are called tensor densities of weight W. Thus an ordinary tensor is a tensor density of weight 0, the most frequently encountered density being of weight $W=1$. To arrive at such a quantity consider the integral

$$\int \mathscr{L}\, \mathrm{d}x^0\, \mathrm{d}x'\, \mathrm{d}x^2\, \mathrm{d}x^3\,. \tag{4.32}$$

In special relativity this integral defines an invariant if $\mathscr{L}$ is a scalar, since under a Lorentz transformation $dx^0 \mathrm{d}x^1 \mathrm{d}x^2 \mathrm{d}x^3 \to \mathrm{d}x'^0 \mathrm{d}x'^1 \mathrm{d}x'^2 \mathrm{d}x'^3$. However, in general relativity for the integral to be an invariant it is necessary that $\mathscr{L}$ be a scalar *density* of weight one. For, consider the transformation to a new co-ordinate system: (*)

$$\mathrm{d}x^0\, \mathrm{d}x^1\, \mathrm{d}x^2\, \mathrm{d}x^3 = \left|\frac{\partial x}{\partial x'}\right| \mathrm{d}x'^0\, \mathrm{d}x'^1\, \mathrm{d}x'^2\, \mathrm{d}x'^3\,, \tag{4.33}$$

and we must have therefore

$$\mathscr{L}' = \left|\frac{\partial x}{\partial x'}\right| \mathscr{L}\,. \tag{4.34}$$

Let us now find an expression for the invariant volume element. We have, since the transformation connecting the metric tensor satisfies

$$g'_{\mu\nu} = \left(\frac{\partial x^{\lambda}}{\partial x'^{\mu}}\right)\left(\frac{\partial x^{\varrho}}{\partial x'^{\nu}}\right) g_{\lambda\varrho}\,, \tag{4.35}$$

and since $g'_{\mu\nu}$ is symmetric we may introduce matrix notation to write the above as

$$\|g'\| = \left\|\frac{\partial x}{\partial x'}\right\|^{T} \|g\| \left\|\frac{\partial x}{\partial x'}\right\|\,, \tag{4.36}$$

where $\|\;\|^{T}$ is the transpose.

(*) Use « outer products » to show this: $\mathrm{d}x^1 \wedge \mathrm{d}x^2 = -\mathrm{d}x^2 \wedge \mathrm{d}x^1$, $\mathrm{d}x^1 \wedge \mathrm{d}x^1 = 0$.

Now using the theorem that the determinant of a product equals the product of the determinants and the fact that the determinant of the transpose matrix equals the determinant of the matrix we have

$$g' = \left|\frac{\partial x}{\partial x'}\right|^2 g\,, \tag{4.37}$$

and hence (*)

$$\sqrt{-g'} = \pm\left|\frac{\partial x}{\partial x'}\right|\sqrt{-g}\,. \tag{4.38}$$

From the above it follows under a co-ordinate transformation

$$\sqrt{-g}\,\mathrm{d}x^0\,\mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3 \Rightarrow \sqrt{-g'}\,\mathrm{d}x'^0\,\mathrm{d}x'^1\,\mathrm{d}x'^2\,\mathrm{d}x'^3\,. \tag{4.39}$$

So that apart from sign $\sqrt{-g}\,\mathrm{d}x^0\mathrm{d}x^1\mathrm{d}x^2\mathrm{d}x^3$ is an invariant element of volume. Because of the $\pm$ sign, however, it is not a true invariant but a « pseudo-invariant ». Another tensor density of importance in four dimensions is the Levi-Civita alternating symbol $e_{\alpha\beta\gamma\delta}$, with $e_{\alpha\beta\gamma\delta}=0$ it two subscripts are equal and $=\pm 1$ depending on whether we have an even or odd permutation $e_{1234}=e_{1234}=-e_{2134}=e_{2143}$. It is also convenient to introduce the pseudotensor $E_{\alpha\beta\gamma\delta}=\sqrt{-g}\,e_{\alpha\beta\gamma\delta}$: provided we deal with proper transformations (for which the Jacobian does not change sign), it transforms as a true tensor. Also we have $E^{\alpha\beta\gamma\delta}=-(-g)^{-\frac{1}{2}}\,e_{\alpha\beta\gamma\delta}$.

4·4. – Covariant differentiation; parallel displacement.

Let us now consider the follow question. Is it possible to introduce derivatives of tensors in a covariant way? The reason for asking this is clear. Our usual laws of physics are formulated in terms of partial differential equations and by our general principle of covariance, these equations should be expressible in a covariant way. If it were not possible to introduce differentiation of tensors in a covariant fashion the theory would be inconsistent. On the other hand we have seen from our discussion of Maxwell's equations in vector form that they are covariant under arbitrary spatial transformations. We therefore surmise that a covariant differentiation formalism should be possible. To arrive at such a formalism it is convenient to consider the problem of differentiating vectors covariantly.

In elementary physics we arrive at the concept of the derivative of a vec-

(*) We introduce the minus sign under the radical since g is negative.

tor $\boldsymbol{V}$ by considering its values at a point x, $\boldsymbol{V}(x)$ and its values at a nearby point $\boldsymbol{V}(x+\mathrm{d}x)$. We then displace the vector parallel to itself and take the difference

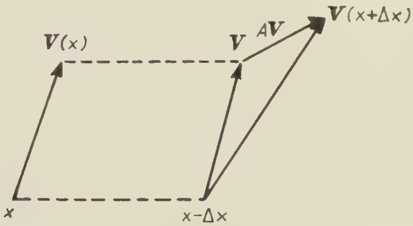


Fig. 2.

so that we have

$$\lim_{\Delta x\to 0}\frac{\boldsymbol{V}(x+\Delta x)-\boldsymbol{V}(x)}{\Delta x}=\frac{\partial \boldsymbol{V}}{\partial x}. \tag{4.40}$$

And in special relativity we arrive at a tensor $T_{ij}=\partial V_i/\partial x^j$ in this fashion. However it is clear that in general $\partial V_\alpha/\partial x^\beta$ does not form a tensor for

$$\frac{\partial V_\alpha}{\partial x^\beta}=\frac{\partial x'^\gamma}{\partial x^\beta}\frac{\partial}{\partial x'^\gamma}\left(\frac{\partial x'^\mu}{\partial x^\alpha}V'_\mu\right)=\frac{\partial x'^\gamma}{\partial x^\beta}\frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial V'_\mu}{\partial x'^\gamma}+\frac{\partial^2 x'^\mu}{\partial x^\beta\,\partial x^\alpha}V'_\mu\,. \tag{4.41}$$

We see that the presence of the second term destroys the transformation properties of $\partial V_\alpha/\partial x^\beta$. Now the interesting thing about this term is that it is proportional to V'_μ and is multiplied by a quantity having three indices $\langle{}^{\mu}_{\beta\alpha}\rangle$. The latter is obviously *not* a tensor since it would vanish under linear transformation of the co-ordinates.

However the appearance of this term suggests that in taking the covariant derivative of V_α we need to add a term linear in V_α which has the effect of cancelling this non-tensor quantity. In other words the covariant derivative will be linear in V_α as we should expect. This additional term represents the change in the vector undergoing parallel displacement from the point x to $x+\Delta x$. We therefore conclude the change in the vector V_α, DV_α, taken between the value at the point $x+\Delta x$ and the value of the vector displaced parallel to itself from the point x will be of the form

$$DV_\alpha=\frac{\partial V_\alpha}{\partial x^\beta}\,\mathrm{d}x^\beta-\Gamma^\mu_{\alpha\beta}V_\mu\,\mathrm{d}x^\beta\,. \tag{4.42}$$

Expressed in derivative form $DV_\alpha/\mathrm{d}x^\beta$ is usually designated as $V_{\alpha;\beta}$ we have

$$V_{\alpha;\beta}=\frac{\partial V_\alpha}{\partial x^\beta}-\Gamma^\mu_{\alpha\beta}V_\mu\,. \tag{4.43}$$

The quantities $\Gamma^{\mu}_{\alpha\beta}$ (yet to be defined) are called the Christoffel symbols of the second kind. From (4.42) it follows that under parallel displacement in the interval $\mathrm{d}x^{\beta}$ a covariant vector undergoes a change

$$\delta V_{\alpha} = + \Gamma^{\mu}_{\alpha\beta} V_{\mu} \mathrm{d}x^{\beta} . \tag{4.44}$$

In analogy with (4.11) we shall assume $\Gamma^{\mu}_{\beta\alpha} = \Gamma^{\mu}_{\alpha\beta}$ although spaces in which this is *not* true have been considered in the literature. To find $V^{\alpha}{}_{;\beta}$ and δV^{α}, we postulate that under parallel displacement the scalar product of two vectors $V^{\alpha}U_{\alpha}$ should be unchanged, as is the case in our usage of parallel displacement in elementary physics. Hence we have (assuming the differentiation law of products)

$$\delta(V^{\alpha}U_{\alpha}) = \delta V^{\alpha} U_{\alpha} + V^{\alpha} \delta U_{\alpha} = 0 . \tag{4.45}$$

Using (4.44) this can be written

$$(\delta V^{\alpha} + \Gamma^{\alpha}_{\mu\beta} V^{\mu} \mathrm{d}x^{\beta}) U_{\alpha} = 0 . \tag{4.46}$$

Since the vector U_{α} is entirely arbitrary, we have

$$\delta V^{\alpha} = - \Gamma^{\alpha}_{\mu\beta} V^{\mu} \mathrm{d}x^{\beta} . \tag{4.47}$$

And it follows

$$V^{\alpha}{}_{;\beta} = \frac{\partial V^{\alpha}}{\partial x^{\beta}} + \Gamma^{\alpha}_{\mu\beta} V^{\mu} . \tag{4.48}$$

Let us now attempt to arrive at an expression for the $\Gamma^{\alpha}_{\mu\beta}$. We have

$$V_{\alpha;\mu} = g_{\alpha\beta;\mu} V^{\beta} + g_{\alpha\beta} V^{\beta}{}_{;\mu} . \tag{4.49}$$

But on the other hand if $V_{\alpha;\mu}$ is to transform as a tensor

$$V_{\alpha;\mu} = g_{\alpha\beta} V^{\beta}{}_{;\mu} \tag{4.50}$$

from which it follows

$$g_{\alpha\beta;\mu} V^{\beta} = 0$$

for all V^{β} and hence,

$$g_{\alpha\beta;\mu} = 0 . \tag{4.51}$$

From (4.48) we therefore have multiplying by $g_{\gamma\alpha}$,

$$V_{\gamma;\beta} = \frac{\partial V_{\gamma}}{\partial x^{\beta}} - \frac{\partial g_{\gamma\alpha}}{\partial x^{\beta}} g^{\nu\alpha} V_{\nu} + \Gamma^{\alpha}_{\mu\beta} g_{\gamma\alpha} g^{\mu\nu} V_{\nu} , \tag{4.52}$$

356

and since $V_{\gamma;\beta} = \partial V_\gamma / \partial x^\beta - \Gamma^\nu_{\gamma\beta} V_\nu$, it follows

$$\Gamma^\nu_{\gamma\beta} = g^{\nu\alpha} \frac{\partial g_{\gamma\alpha}}{\partial x^\beta} - \Gamma^\alpha_{\mu\beta} g_{\gamma\alpha} g^{\mu\nu} . \tag{4.53}$$

Let us now introduce the Christoffel symbols of the first kind defined by

$$\Gamma^\mu_{\gamma\beta} = g^{\mu\alpha}[\alpha, \gamma\beta] . \tag{4.54}$$

Then (4.53) becomes

$$[\varrho, \gamma\beta] = \frac{\partial g_{\varrho\gamma}}{\partial x^\beta} - [\gamma, \varrho\beta] , \tag{4.55}$$

which may be written upon permuting the indices in the following ways,

$$\left\{ \begin{aligned} [\varrho, \gamma\beta] + [\gamma, \varrho\beta] &= \frac{\partial g_{\varrho\gamma}}{\partial x^\beta} , \\ [\gamma, \beta\varrho] + [\beta, \gamma\varrho] &= \frac{\partial g_{\gamma\beta}}{\partial x^\varrho} , \\ [\beta, \varrho\gamma] + [\varrho, \beta\gamma] &= \frac{\partial g_{\beta\varrho}}{\partial x^\gamma} . \end{aligned} \right. \tag{4.56}$$

Now using the symmetry $[\beta, \gamma\varrho] = [\beta, \varrho\gamma]$ and subtracting the 2nd equation from the 3rd we have

$$- [\gamma, \beta\varrho] + [\varrho, \beta\gamma] = \frac{\partial g_{\varrho\beta}}{\partial x^\gamma} - \frac{\partial g_{\gamma\beta}}{\partial x^\varrho} . \tag{4.57}$$

And now upon adding to the first we find

$$[\varrho, \beta\gamma] = \frac{1}{2}\left(\frac{\partial g_{\varrho\gamma}}{\partial x^\beta} + \frac{\partial g_{\varrho\beta}}{\partial x^\gamma} - \frac{\partial g_{\beta\gamma}}{\partial x^\varrho} \right) . \tag{4.58}$$

And hence upon multiplication by $g^{\nu\varrho}$

$$\Gamma^\nu_{\beta\gamma} = \frac{1}{2} g^{\nu\varrho} \left(\frac{\partial g_{\varrho\gamma}}{\partial x^\beta} + \frac{\partial g_{\varrho\beta}}{\partial x^\gamma} - \frac{\partial g_{\beta\gamma}}{\partial x^\varrho} \right) . \tag{4.59}$$

Finally we may derive the rules for differentiation of higher order tensors by treating them as products, for example, write $T_{\mu\nu} = A_\mu B_\nu$; then

$$T_{\mu\nu;\lambda} = A_{\mu;\lambda} B_\nu + A_\mu B_{\nu;\lambda} = \frac{\partial T_{\mu\nu}}{\partial x^\lambda} - \Gamma^\varrho_{\mu\lambda} T_{\varrho\nu} - \Gamma^\zeta_{\nu\lambda} T_{\mu\zeta} . \tag{4.60}$$

Using the rules for covariant differentiation one may recast Maxwell's equations in four dimensional covariant form. They become

$$F_{\mu\nu;\lambda}+F_{\nu\lambda;\mu}+F_{\lambda\mu;\nu}=0\,,\qquad \mathscr{F}^{\mu\nu}{}_{;\nu}=J^{\mu}\,, \tag{4.61}$$

where $\mathscr{F}^{\mu\nu}=\sqrt{-g}\,F^{\mu\nu}$, $J^{\mu}=\sqrt{-g}\,j^{\mu}$. It is a curious result that because $F^{\mu\nu}=-F^{\nu\mu}$ all the above covariant derivatives may be replaced by ordinary derivatives. The « dual » to the $F^{\mu\nu}$ is defined by $F^{*}_{\lambda\varrho}=\frac{1}{2}E_{\lambda\varrho\mu\nu}F^{\mu\nu}$; show that the dual also satisfies the homogeneous Maxwell equations. Note that we have $F^{\mu\nu}=-\frac{1}{2}E^{\mu\nu\lambda\varrho}F^{*}_{\lambda\varrho}$. In differentiating tensor densities the following rule is necessary, which can be easily derived using $g_{\alpha\beta;\lambda}=0$, $g^{\alpha\beta}{}_{;\lambda}=0$ and the definition of the determinant, *i.e.*, $g_{;\lambda}=0$. Also one has $\Gamma^{\alpha}_{\beta\alpha}=\frac{1}{2}g^{\mu\alpha}\,\partial g_{\mu\alpha}/\partial x^{\beta}=\partial\ln\sqrt{-g}/\partial x^{\beta}$. Using this we find for a symmetric tensor

$$T^{\mu}_{\nu;\mu}=\frac{1}{\sqrt{-g}}\,\frac{\partial\sqrt{-g}\,T^{\mu}_{\nu}}{\partial x^{\mu}}-\frac{1}{2}\,\frac{\partial g_{\lambda\varrho}}{\partial x^{\nu}}\,T^{\lambda\varrho}\,. \tag{4.62}$$

4·5. – The equations of motion of a particle; geodesic paths.

The concepts of parallel displacement of a vector and covariant differentiation may be used to formulate the equations of geodesic motion in a very succint way. In classical mechanics we have the well-known variational principle

$$\delta\int_{t_1}^{t_2}L\,\mathrm{d}t=0\,, \tag{4.63}$$

where L is the Lagrangian and is generally of the form $L(q,\dot{q})$, where q, $\dot{q}$ are the generalized co-ordinates and velocities, and t_1 and t_2 are the points between which the variation is taken.

In general relativity we adopt a similar variational principle which reduces to the above in the non-relativistic limit,

$$\delta\int_{P_1}^{P_2}\mathrm{d}s=\delta\int_{P_1}^{P_2}\sqrt{g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu}}=0\,. \tag{4.64}$$

The mathematical interpretation of this being that we are looking for those paths between the space-time points P_1 and P_2 for which the four dimen-

sional distance is an extremum and the integral stationary. Such paths are called geodesic paths. In the case of special relativity the $g_{\mu\nu}$ may be taken to be constants and in this case the path becomes simply an Euclidean straight line. However in general this is not the case. A simple physical interpretation of the variational principle is to note that since ds represents the proper interval of a clock, we are simply comparing the readings of clocks moved between P_1 and P_2. From special relativity it is clear that if we move a clock C_1 uniformly between two space points A and B, and if we move another clock C_2 along a winding path between A and B, and if moreover we demand that C_2 left A and arrived at B simultaneously with C_1, then it follows, C_2 will read *less* elapsed time than C_1. Thus in special relativity the geodesic path will make $\int_{P_1}^{P_2} \mathrm{d}s$ a *maximum* compared with neighboring paths, although more generally all we can say is that it will be an extremum.

Let us now introduce some parameter λ so that the paths are describable as $x^\mu = x^\mu(\lambda)$ and we may write (4.64) as

$$\delta \int_{P_1(\lambda_1)}^{P_2(\lambda_2)} \sqrt{g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}}\, \mathrm{d}\lambda = 0 . \tag{4.65}$$

In this form we now have a typical variational principle with fixed end points. Therefore designating the integrand by $L(q, \dot{q})$ where $\dot{q} = \mathrm{d}x^\mu/\mathrm{d}\lambda$ we arrive at the Euler-Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\partial L}{\partial \dot{x}^\varrho} - \frac{\partial L}{\partial x^\varrho} = 0 . \tag{4.66}$$

The expressions $\partial L/\partial x$, $\partial L/\partial \dot{x}^\varrho$ may be written

$$\frac{\partial L}{\partial x^\varrho} = \frac{1}{2} \frac{1}{L} \frac{\partial g_{\mu\nu}}{\partial x^\varrho} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} , \tag{4.67}$$

$$\frac{\partial L}{\partial \dot{x}^\varrho} = \frac{1}{2} \frac{1}{L} g_{\varrho\nu} \dot{x}^\nu + \frac{1}{2L} g_{\mu\varrho} \dot{x}^\mu = \frac{1}{L} g_{\varrho\nu} \dot{x}^\nu . \tag{4.68}$$

and hence

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \frac{\partial L}{\partial \dot{x}^\varrho} = - \frac{1}{L^2} \frac{\mathrm{d}L}{\mathrm{d}\lambda} g_{\varrho\nu} \dot{x}^\nu + \frac{1}{L} \frac{\partial g_{\varrho\nu}}{\partial x^\mu} \dot{x}^\mu \dot{x}^\nu + g_{\varrho\nu} \ddot{x}^\nu . \tag{4.69}$$

Now let us choose $L = \mathrm{d}s/\mathrm{d}\lambda = 1$. This means we are taking as our para-

meter the element of the distance itself. This can be done in general except for a light ray along which $ds = 0$; for this case one has to resort to a limiting process. Then we have with $L = 1$,

$$g_{\varrho\nu}\frac{d^2x^\nu}{ds^2} + \frac{\partial g_{\varrho\nu}}{\partial x^\mu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\varrho}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0\,. \tag{4.70}$$

By interchanging the « dummy indices » in the second term we may write it as

$$\frac{\partial g_{\varrho\nu}}{\partial x^\mu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = \frac{1}{2}\left(\frac{\partial g_{\varrho\nu}}{\partial x^\mu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} + \frac{\partial g_{\varrho\mu}}{\partial x^\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}\right). \tag{4.71}$$

and hence finally the above can be written

$$g_{\varrho\nu}\frac{d^2x^\nu}{ds^2} + [\varrho, \mu\nu]\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0\,. \tag{4.72}$$

Now multiplying by $g^{\lambda\varrho}$ and using $g^{\lambda\varrho}g_{\varrho\nu} = \delta^\lambda_\nu$, $g^{\alpha\varrho}[\varrho, \mu\nu] = \Gamma^\lambda_{\mu\nu}$ we have finally

$$\frac{d^2x^\alpha}{ds^2} + \Gamma^\alpha_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0\,. \tag{4.73}$$

The equations are the generalization in the presence of a gravitational field of the special relativistic equations

$$\frac{d^2x^\alpha}{ds^2} = 0\,. \tag{4.74}$$

The term $\Gamma^\lambda_{\mu\nu}(dx^\mu/ds)(dx^\nu/ds)$ represents the effect of the gravitational field although in flat space the $\Gamma^\alpha_{\mu\nu}$ will not in general vanish for arbitrary choices of the co-ordinates. Because the $\Gamma^\alpha_{\mu\nu}$ are not the components of a tensor and depend only on the first derivatives of the $g_{\mu\nu}$ it is always possible to choose a co-ordinate system such that *locally* the $\Gamma^\alpha_{\mu\nu}$ vanish. Having done this the equations of motion reduce to those of special relativity. To arrive at this result we note that it can be shown by *direct calculation* that the transformation properties of the $\Gamma^\alpha_{\mu\nu}$ are

$$\Gamma'^\alpha_{\mu\nu} = \frac{\partial x'^\alpha}{\partial x^\beta}\frac{\partial x^\gamma}{\partial x'^\mu}\frac{\partial x^\delta}{\partial x'^\nu}\Gamma^\beta_{\gamma\delta} + \frac{\partial x'^\alpha}{\partial x^\varrho}\frac{\partial^2 x^\varrho}{\partial x'^\mu \partial x'^\nu}\,. \tag{4.75}$$

Now let us choose the transformation to be of the form

$$x^\varrho = x'^\varrho + \tfrac{1}{2}a^\varrho_{\mu\nu}x'^\mu x'^\nu \tag{4.76}$$

in the neighborhood of the point $x'^{\varrho} = 0$. Then upon performing the differentiation we have

$$\Gamma'^{\alpha}_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu} + a^{\alpha}_{\mu\nu} = 0 \tag{4.77}$$

and hence choosing $a^{\alpha}_{\mu\nu} = -\Gamma^{\alpha}_{\mu\nu}$ the result follows. The system of co-ordinates chosen in this way is called a geodesic system or geodesic frame. In such a frame all the derivatives $\partial g_{\mu\nu}/\partial x'^{\lambda}$ vanish locally. There are forty such derivatives and similarly 40 Christoffel symbols and the two sets of quantities are linearly dependent. Thus we have

$$\begin{cases} 0 = g_{\mu\nu;\lambda} = \dfrac{\partial g_{\mu\nu}}{\partial x^{\lambda}} - \Gamma^{\alpha}_{\mu\nu} g_{\alpha\nu} - \Gamma^{\beta}_{\nu\lambda} g_{\mu\beta}\,. \\ \dfrac{\partial g_{\mu\nu}}{\partial x^{\lambda}} = \Gamma^{\alpha}_{\mu\lambda} g_{\alpha\nu} + \Gamma^{\beta}_{\nu\lambda} g_{\mu\beta}\,. \end{cases} \tag{4.78}$$

Hence the vanishing of the Γ's implies the vanishing of the $\partial g_{\mu\nu}/\partial x^{\lambda}$ and conversely. In such a frame the gradients of the gravitational potentials vanish and consequently the « intensities » or « field strengths » of the gravitational field. Finally, we note a more general theorem due to FERMI (1922), namely, we may choose co-ordinates such that they vanish along a geodesic path, *i.e.*, not merely at a given point. Einsteins discussions of observations in a freely-falling elevator, together with many other discussions preceding Fermi's work, appear to implicitly assume this theorem. It may therefore be regarded as a logical « check » of the theory that it implicitly « predicted » the existence of such a theorem.

4·6. – Riemann curvature tensor.

We have now seen how it is possible to develop a formalism to handle the problem of covariant differentiation in spaces for which the $g_{\mu\nu}$ are not constants, treating the time co-ordinate on the same footing as the spatial co-ordinates. In the case of spaces in which it is possible to introduce Cartesian co-ordinates the covariant derivative reduces to the ordinary derivative. However, for some spaces it is not possible to do this: for example, the surface of a sphere forms a two dimensional « curved space ». It is not possible to find a co-ordinate system for which the element of distance takes the simple form $d\sigma^2 = dx^2 + dy^2$ and the g_{ij} $(i, j = 1, 2)$ all become constants. For the sphere this is essentially intuitively obvious, however for more complicated spaces we clearly need some mathematical formalism for determining whether or not a space is curved. The physical application will be to associate certain types of curved spaces with the presence of permanent gravitational fields such as

those produced by gravitating bodies as the earth, the sun, and in fact any distribution of energy (by the principle of equivalence). When there are no such bodies present, at least locally, we may imagine that their effect on the curvature becomes negligible, and hence by a suitable co-ordinate transformation the space may be made flat (*).

We therefore proceed to look for a tensor that will characterize the curvature of space. The reason we look for a tensor is clear: if we have a tensor and it vanishes in one co-ordinate system it vanishes in all co-ordinate systems, so that our characterization of whether or not a space is flat will not depend on the particular system of co-ordinates chosen. Conversely if the space is not flat and hence the tensor is not zero, then it will not vanish in any other co-ordinate system we may adopt in the space.

The above properties of a tensor are very fundamental and may be proved simply as follows; we have for an arbitrary tensor T^{μ}_{ν}

$$T'^{\mu\dots}_{\nu} = \left(\frac{\partial x'^{\mu}}{\partial x^{\lambda}}\right)\left(\frac{\partial x^{\varrho}}{\partial x'^{\nu}}\right) \dots T^{\lambda\dots}_{\varrho\dots}\,. \tag{4.79}$$

Then if $T^{\lambda\dots}_{\varrho\dots}$ vanishes, clearly $T'^{\mu\dots}_{\nu\dots}$ also vanishes. On the other hand if $T^{\lambda\dots}_{\varrho\dots}$ does not vanish, neither can the $T'^{\mu\dots}_{\nu\dots}$ vanish. For we have also

$$T^{\lambda\dots}_{\varrho\dots} = \left(\frac{\partial x^{\lambda}}{\partial x'^{\mu}}\right)\left(\frac{\partial x'^{\nu}}{\partial x^{\varrho}}\right) \dots T'^{\mu\dots}_{\nu\dots}\,. \tag{4.80}$$

and if $T'^{\mu\dots}_{\nu\dots}$ vanished so would $T^{\lambda\dots}_{\varrho\dots}$, contrary to our assumption.

To find such a tensor it is convenient to consider what happens to a vector when we transport it parallel to itself around a closed circuit. We know in flat space when we transport a vector parallel to itself around such a circuit the vector is unchanged; indeed this is one of the basic assumptions of Euclidean geometry and consequently of elementary mechanics as well (the parallelogram of forces).

However consider the surface of a sphere and consider a unit vector $dx^i/d\sigma$ tangent to a geodesic on the sphere (an arc of a great circle), then on parallel displacement, by assumption a vector V^i will maintain a constant angle with

---

(*) No attempt will be made here to recount the exciting history of the development of curved spaces and non-Euclidean geometry. The most important contributions are associated with the names of SACCHERI, LEGENDRE, GAUSS, LOBATCHEVSKY, BOLYAI, RIEMANN and BELTRAMI (who proved the consistency of Euclidean geometry implies the consistency of non-Euclidean geometry of the hyperbolic type). For a history of this development see, for example R. BONOLA: *Non-Euclidean Geometry*.

respect to $dx^i/d\sigma$. As can be easily seen from the figure upon displacing V^i parallel to itself around the circuit $A\ B\ C\ A$ it will, upon returning to A, no longer point in the original direction. Clearly had we chosen another circuit we would in general find a different discrepancy. In other words, in general, in a curved space if we transport a vector parallel to itself along two different paths from a point P to a point P' the two vectors will in general have different components at P'.

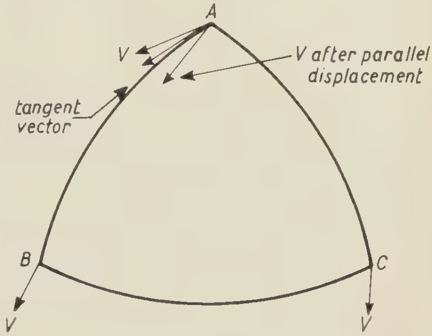


Fig. 3.

We now proceed to utilize this idea to construct the curvature tensor. We consider the infinitesimal parallelogram of sides δx^λ, Δx^ϱ and a vector V^μ (*). We transport V^μ parallel to itself from the point (1) to the point (3) first along the path 123, and then along the path 143. The difference in the two vectors will then be a measure of the curvature of the space. We therefore have upon transporting the vector $V^\mu(1)$ parallel to itself to the point (2), since $\delta V^\mu(1 \to 2) = -\Gamma^\mu_{\lambda\nu}(1)\,\delta x^\lambda V^\nu(1)$

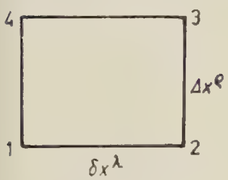


Fig. 4.

$$V^\mu(2) = V^\mu(1) - \Gamma^\mu_{\lambda\nu}(1)\,\delta x^\lambda V^\nu(1)\,. \tag{4.81}$$

Upon transporting it to (3), we must take into account the change undergone in $\Gamma^\mu_{\lambda\nu}$; we have

$$\delta V^\mu(2 \to 3) = -\left\{\Gamma^\mu_{\varrho\nu}(1) + \frac{\partial \Gamma^\mu_{\varrho\nu}(1)}{\partial x^\lambda}\,\delta x^\lambda\right\} V^\nu(2)\,\Delta x^\varrho\,, \tag{4.82}$$

and substituting for $V^\nu(2)$ and neglecting terms of higher order than $\delta x^\lambda \Delta x^\varrho$ we have

$$\delta V^\mu(2 \to 3) = -\Gamma^\mu_{\varrho\nu} V^\nu(1)\,\Delta x^\varrho + \Gamma^\mu_{\varrho\nu}(1)\,\Gamma^\nu_{\lambda\alpha}(1)\,V^\alpha(1)\,\Delta x^\varrho\,\delta x^\lambda - \frac{\partial \Gamma^\mu_{\varrho\nu}(1)}{\partial x^\lambda}\,V^\nu(1)\,\delta x^\lambda \Delta x^\varrho\,. \tag{4.83}$$

and the value of the vector at the point (3) is

$$\underset{(123)}{V^\mu}(3) = V^\mu(1) - \Gamma^\mu_{\lambda\nu}(1)\,\delta x^\lambda V^\nu(1) + \delta V^\mu(2 \to 3)\,. \tag{4.84}$$

(*) Strictly speaking the « parallelogram » does not close up unless we modify either δx^λ, or Δx^λ or both; however, the effects we are neglecting are of third order in the infinitesimals, whereas the change in V^μ for which we are looking is second order. See, *e.g.*, E. CARTAN: *Leçons sur la géométrie des espaces de Riemann* (Paris, 1951).

On the other hand if we transport the vector along the path (143) we have

$$(4.85)\qquad \underset{(143)}{V^\mu}(3) = V^\mu(1) - \Gamma^\mu_{\varrho\nu}(1)\,\Delta x^\varrho V^\nu(1) + \delta V^\mu(3\to 4)\,,$$

where $\delta V^\mu(3\to 4)$ is given by

$$(4.86)\qquad \delta V^\mu(3\to 4) = -\,\Gamma^\mu_{\lambda\nu}V^\nu(1)\,\delta x^\lambda + \Gamma^\mu_{\lambda\nu}(1)\,\Gamma^\nu_{\varrho\alpha}(1)\,V^\alpha(1)\,\Delta x^\varrho\delta x^\lambda - \frac{\partial\Gamma^\mu_{\lambda\nu}(1)}{\partial x^\varrho}\,V^\nu(1)\,\delta x^\lambda\Delta x^\varrho\,.$$

Upon taking the difference $\underset{(123)}{V^\mu}(3) - \underset{(143)}{V^\mu}(3)$

$$(4.87)\qquad \underset{(123)}{V^\mu}(3) - \underset{(143)}{V^\mu}(3) = \left(\frac{\partial\Gamma^\mu_{\alpha\lambda}}{\partial x^\varrho} - \frac{\partial\Gamma^\mu_{\alpha\varrho}}{\partial x^\lambda} - \Gamma^\mu_{\lambda\nu}\Gamma^\nu_{\varrho\alpha} + \Gamma^\mu_{\varrho\nu}\Gamma^\nu_{\lambda\alpha}\right)V^\alpha\Delta x^\varrho\,\delta x^\lambda.$$

Note that the quantity $\underset{(123)}{V^\mu}(3) - \underset{(143)}{V^\mu}(3)$ is also the change δV^μ upon taking V^μ around the closed circuit, since we have $\underset{(123)}{V^\mu}(3) = V^\mu(1) + \delta\underset{(123)}{V^\mu}$, $\underset{(143)}{V^\mu}(3) = V^\mu(1) + \delta\underset{(143)}{V^\mu}$, and hence, taking the difference $\delta\underset{(123)}{V^\mu} - \delta\underset{(143)}{V^\mu} = \delta\underset{(1234)}{V^\mu}$. The negative of the quantity in parentheses is called the Riemann curvature tensor and is denoted by $R^\mu{}_{\alpha\lambda\varrho}$ (although other notations are frequently used in the literature). We may therefore rewrite (4.87) as

$$(4.88)\qquad \delta\underset{(1234)}{V^\mu} = -\,R^\mu{}_{\alpha\lambda\varrho}V^\alpha\,\Delta x^\varrho\,\delta x^\lambda\,.$$

Upon observing that $R^\mu{}_{\alpha\lambda\varrho} = -\,R^\mu{}_{\alpha\varrho\lambda}$ we may further rewrite the expression as

$$(4.89)\qquad \delta\underset{(1234)}{V^\mu} = \tfrac{1}{2}R^\mu{}_{\alpha\lambda\varrho}V^\alpha\,\Delta\sigma^{\lambda\varrho}\,,$$

where $\Delta\sigma^{\lambda\varrho} = \Delta x^\lambda\,\delta x^\varrho - \Delta x^\varrho\,\delta x^\lambda$ and represents the infinitesimal element of area. Since V^α, $\Delta\sigma^{\lambda\varrho}$ and $\delta\underset{(1234)}{V^\mu}$ are tensors, it is clear that $R^\mu{}_{\alpha\lambda\varrho}$ must also be a tensor. The reason that $\delta\underset{(1234)}{V^\mu}$ transforms as a tensor is due to the fact that it represents the comparison of two vectors (the original vector and the vector transported around the closed path) at a point and this clearly defines another vector at the point.

Another way to see that $R^\mu{}_{\alpha\lambda\varrho}$ is a tensor is to directly compute $V^\mu{}_{;\lambda;\varrho} - V^\mu{}_{;\varrho;\lambda}$, one finds

$$(4.90)\qquad V^\mu{}_{;\lambda;\varrho} - V^\mu{}_{;\varrho;\lambda} = -\,R^\mu{}_{\alpha\lambda\varrho}V^\alpha$$

from which the tensorial character of $R^\mu{}_{\alpha\lambda\varrho}$ is clear.

Thus the non-vanishing of the Riemann curvature tensor introduces a non-commutativity of covariant derivatives. The vanishing of $R^\mu{}_{\alpha\lambda\varrho}$ is clearly a *necessary* condition that the space be flat, it may also be shown that it is a *sufficient* condition.

394

4·7. – Symmetry properties of the Riemann tensor and Bianchi identities.

Another expression for the Riemann tensor is given by

$$R_{\lambda\mu\nu\varrho} = g_{\lambda\varkappa} R^{\varkappa}{}_{\mu\nu\varrho} . \tag{4.91}$$

Upon introducing the Christoffel symbols of the first kind and denoting ordinary differentiation by a comma, the above expression may be written

$$R_{\lambda\mu\nu\varrho} = [\lambda, \mu\varrho]_{,\nu} - [\lambda, \mu\nu]_{,\varrho} + g^{\varkappa\alpha}\{[\varkappa, \lambda\varrho][\alpha, \mu\nu] - [\varkappa, \lambda\nu][\alpha, \mu\varrho]\} . \tag{4.92}$$

The first two terms may be rewritten,

$$[\lambda, \mu\varrho]_{,\nu} - [\lambda, \mu\nu]_{,\varrho} = \tfrac{1}{2}(g_{\lambda\varrho,\mu\nu} + g_{\mu\nu,\lambda\varrho} - g_{\lambda\nu,\mu\varrho} - g_{\mu\varrho,\lambda\nu}) . \tag{4.93}$$

From these expressions it may be verified that $R_{\lambda\mu\nu\varrho}$ has the following symmetry properties:

$$\begin{cases} R_{\lambda\mu\nu\varrho} = -R_{\lambda\mu\varrho\nu} \\ R_{\lambda\mu\nu\varrho} = +R_{\nu\varrho\lambda\mu} \\ R_{\lambda\mu\nu\varrho} = -R_{\mu\lambda\nu\varrho} \end{cases} \tag{4.94}$$

$$R_{\lambda\mu\nu\varrho} + R_{\lambda\nu\varrho\mu} + R_{\lambda\varrho\mu\nu} = 0 \qquad \text{(cyclic identities)} . \tag{4.95}$$

In addition to the above identities we have also some identities involving the first derivatives of the $R_{\lambda\mu\nu\varrho}$, the « Bianchi identities », which state

$$R_{\lambda\mu\nu\varrho;\zeta} + R_{\lambda\mu\varrho\zeta;\nu} + R_{\lambda\mu\zeta\nu;\varrho} = 0 . \tag{4.96}$$

These identities may be readily verified upon adopting geodesic co-ordinates at a given point so that the non-linear terms in (4.92) vanish and using (4.93).

As a consequence of the identities (4.94), (4.95) the number of independent components N_I of the Riemann tensor is considerably reduced from the n^4 (where $n = \delta^\mu_\mu$, the dimensionality of the space) terms one would otherwise have. To arrive at an expression for N_I in terms of n we note that there are three cases to consider: 1) case where two subscripts are repeated, *e.g.*, terms like R_{1212}; 2) case where one subscript is repeated, *e.g.*, R_{1213}; 3) case where no subscripts are repeated. In case 1), we have $\frac{1}{2}n(n-1)$ independent terms from (4.94), moreover (4.95) yields no reduction, since terms like R_{abab} upon applying the cyclic identities yield $R_{abab} + R_{aabb} + R_{abba} = 0$ but $R_{aabb} = 0$ from (4.94) and hence the identity merely reproduces what we have already used

$R_{abab} = -R_{abba}$. (We have used latin letters above to indicate no summation is intended). In case 2) we have $\frac{1}{2}n(n-1)(n-2)$ possibilities, also from $R_{abac} + R_{aacb} + R_{acba} = 0$, the cyclic identities do not reduce this number. In case 3), neglecting the cyclic identities, we have $\frac{1}{2}\cdot\frac{1}{2}n(n-1)\frac{1}{2}(n-2)(n-3)$ possibilities, the extra factor $\frac{1}{2}$ arising because of the symmetry $R_{\lambda\mu\nu\varrho} = R_{\nu\varrho\lambda\mu}$. Because of the cyclic identities this number is reduced by the factor $\frac{2}{3}$. Hence we have in total

$$N_I = \frac{1}{2}n(n-1) + \frac{1}{2}n(n-1)(n-2) + \frac{2}{3}\frac{1}{2}\frac{n(n-1)}{2}\frac{(n-2)(n-3)}{2} = \frac{1}{12}n^2(n^2-1)\,. \tag{4.97}$$

It is interesting to tabulate for a range of n, the values of N_I, and also the number of components of the $g_{\mu\nu} = \frac{1}{2}n(n+1)$ and also of the Christoffel symbols, $\frac{1}{2}n^2(n+1)$, we have

(4.98)

n	$\frac{1}{2}n(n+1)$	$\frac{1}{2}n^2(n+1)$	N_I
2	3	6	1
3	6	18	6
4	10	40	20
5	15	75	50
6	21	126	105
7	28	196	196
8	36	288	336

We see, for example, that the two-dimensional manifold is particularly simple since its curvature tensor has only one independent component, which we make take to be R_{1212}.

4·8. – Contracted curvature tensor; the Ricci tensor.

We may contract the Riemann tensor to form a tensor of rank two, the Ricci tensor, defined as

$$R_{\mu\nu} = g^{\lambda\varrho}R_{\lambda\mu\nu\varrho}\,. \tag{4.99}$$

The tensor $R_{\mu\nu}$ is symmetric since using (4.94)

$$R_{\mu\nu} = g^{\lambda\varrho}R_{\lambda\mu\nu\varrho} = g^{\lambda\varrho}R_{\nu\varrho\lambda\mu} = g^{\lambda\varrho}R_{\varrho\nu\mu\lambda} \tag{4.100}$$

336

we may further contract to form the scalar curvature R

$$R \equiv R^{\mu}_{\mu} = g^{\mu\nu} g^{\lambda\varrho} R_{\lambda\mu\nu\varrho} \,. \tag{4.101}$$

If we now consider the case of $n=2$, we find for R the following expression

$$R = -\frac{2R_{1212}}{g} \,. \tag{4.102}$$

The quantity R_{1212}/g is called the Gaussian curvature K of the space

4.103)
$$K = \frac{R_{1212}}{g} \,.$$

If we calculate K for the surface of a sphere with the line element given by

$$\mathrm{d}\sigma^2 = R^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2) = \frac{\mathrm{d}r^2}{1 - r^2/R^2} + r^2\,\mathrm{d}\varphi^2 \,. \qquad r = R\sin\theta \,. \tag{4.104}$$

we find $K = R^{-2}$. Similarly for the pseudo-sphere obtained by setting $R \to iR$, $\theta \to -i\theta'$, we find $K = -R^{-2}$. (Calculate the Gaussian curvature for a two-dimensional surface associated with the spatial line-element of Chapter II (eq. (2.11)) in the « plane » $\mathrm{d}z = 0$.)

The geometrical significance of the Ricci tensor is obtained by calculating the sum of the Gaussian curvatures for the set of orthogonal planes spanned by geodesics through a given point in the n-dimensional Riemannian manifold (alternatively referred to as « V_n »). Denoting the curvature associated with a given plane by $K_{\alpha\beta}$, one finds (*)

(4.10)
$$\sum_{\beta} K_{\alpha\beta} = - R_{\mu\nu} \lambda^{\mu}_{(\alpha)} \lambda^{\nu}_{(\alpha)} \,,$$

where $\lambda^{\mu}_{(\alpha)}$ is a unit vector. The sum of these mean curvatures in turn defines the negative of the scalar curvature, using $\sum_{\alpha} \lambda^{\mu}_{(\alpha)} \lambda^{\nu}_{(\alpha)} = g^{\mu\nu}$.

(*) See, for example, C. E. Weatherburn: *An Introduction to Riemannian Geometry and the Tensor Calculus* (Cambridge, 1938); L. P. Eisenhart: *Riemannian Geometry* (Princeton, 1926); T. Levi-Civita: *The Absolute Differential Calculus* (London, 1927).

397

Chapter V.

THE FIELD EQUATIONS AND SCHWARZSCHILD'S SOLUTION; EQUATIONS OF MOTION IN THE SCHWARZSCHILD FIELD

5·1. – The Einstein tensor.

In the preceding chapter we arrived at the Riemann curvature tensor $R^\lambda{}_{\mu\nu\varrho}$ which in a V_4 has 20 independent components. One might now consider constructing a theory of gravitation based directly on $R^\lambda{}_{\mu\nu\varrho}$. However, as we intend to associate the presence of a permanent gravitational field with material sources (just as we associate an electromagnetic field with charged sources, it would be necessary to relate $R^\lambda{}_{\mu\nu\varrho}$ to a tensor of the same rank representing a description of matter, and we do not have a 20 component tensor providing such a description. Moreover we have only 10 field quantities, the $g_{\mu\nu}$, and so we would also expect that actually we need only 10 equations. Further, outside a material source, we intend to assert that the space-time manifold is curved, and clearly if our equations were of the form

$$R^\lambda{}_{\mu\nu\varrho} \propto M^\lambda{}_{\mu\nu\varrho} , \tag{5.1}$$

where $M^\lambda{}_{\mu\nu\varrho}$ described the matter-source, then we should be forced to conclude that outside matter, space would be flat. The geodesic paths would then be straight lines and we would no longer have a possible geometrical explanation for the phenomena associated with permanent gravitational fields—in contradiction with our physical arguments based on the principle of equivalence. We therefore rule out any attempt to construct field equations on the basis of (5.1).

Thus we are left with $R_{\mu\nu}$, R as the simplest possibilities. A theory based on

$$R \propto T , \tag{5.2}$$

where $T \equiv T^\mu_\mu$ (= trace of the energy-momentum tensor) is also ruled-out on the grounds that it does not give us enough equations to determine the $g_{\mu\nu}$. We note that because of the possibility of performing four co-ordinate transformations we can in general make four of the $g_{\mu\nu}$ constant, thus leaving us with six functions to determine. From (5.2) we could eliminate one, but the system is hopelessly undetermined. Thus a theory based on (5.2) while more in line with what we expect, is still not satisfactory.

398

We are therefore forced by these arguments to attempt to construct a theory based on $R_{\mu\nu}$, and $T_{\mu\nu}$ the symmetric energy-momentum tensor. A natural starting point, and indeed Einstein's original starting point was to assume

$$R_{\mu\nu} \propto T_{\mu\nu} . \tag{5.3}$$

However, this led to the following difficulty: We know from special relativity that the $T_{\mu\nu}$ satisfy certain conservation laws which in covariant notation are $T^{\mu}_{\nu;\mu} = 0$; on the other hand, one may readily show $R^{\mu}_{\nu;\mu} \neq 0$, unless $R = 0$. For a while EINSTEIN thought that (5.3) was nevertheless correct and since (5.3) implies $R \propto T$, that T must vanish for matter just as it does for the electromagnetic field. However this led to serious difficulties; moreover in analogy with Maxwell's equations it would be more satisfactory if one could obtain $T^{\mu}_{\nu;\mu} = 0$ because of an identity in the field equations. In this way one would also avoid having to make specializing asumptions about the $T_{\mu\nu}$.

The fact that we need a differential identity suggests studying the Bianchi identities. Let us therefore contract the Bianchi identities

$$R^{\alpha}{}_{\beta\gamma\delta;\nu} + R^{\alpha}{}_{\beta\delta\nu;\gamma} + R^{\alpha}{}_{\beta\nu\gamma;\delta} = 0 ,$$

with respect to the indices α and δ. We have

$$R_{\beta\gamma;\nu} - R_{\beta\nu;\gamma} + R^{\alpha}{}_{\beta\nu\gamma;\alpha} = 0 . \tag{5.4}$$

Upon raising the index β, we have

$$R^{\beta}{}_{\gamma;\nu} - R^{\beta}{}_{\nu;\gamma} + R^{\alpha\beta}{}_{\nu\gamma;\alpha} = 0 . \tag{5.5}$$

Now contracting with respect to β and ν this becomes

$$2R^{\beta}{}_{\gamma;\beta} - R_{;\gamma} = 0 \tag{5.6}$$

which may be alternatively written on using $\delta^{\beta}_{\gamma} R_{;\beta} = R_{;\gamma}$

$$(R^{\beta}_{\gamma} - \tfrac{1}{2}\delta^{\beta}_{\gamma} R)_{;\beta} = 0 . \tag{5.7}$$

The tensor defined by

$$G^{\mu}_{\nu} = R^{\mu}_{\nu} - \tfrac{1}{2}\delta^{\mu}_{\nu} R \tag{5.8}$$

has the properties of the tensor for which we were looking. Lowering the index μ we have alternatively $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$. The tensor $G_{\mu\nu}$ is called the

Einstein tensor. Unfortunately the tensor $G_{\mu\nu}$ is not unique; we could have added a tensor with vanishing divergence, in particular the cosmological term $\Lambda g_{\mu\nu}$, since $g_{\mu\nu;\lambda}=0$. The problem of whether this term does or does not belong in the equations has never been settled, the prevailing opinion being that it does not belong in the equations—in any case the « cosmological constant » Λ must be small, of order $10^{-56}\ \text{cm}^{-2}$, and would not lead to any measurable effects on planetary dynamics. We shall therefore disregard it in this chapter.

Before proceeding to construct the field equations let us now obtain explicit expressions for the Ricci tensor. We find upon using the identity $\Gamma^{\lambda}_{\beta\lambda}=\partial_\beta \ln\sqrt{-g}$,

$$(5.9)\qquad R_{\mu\nu}=\frac{\partial^2\ln\sqrt{-g}}{\partial x^\mu\,\partial x^\nu}-\frac{\partial\Gamma^\alpha_{\mu\nu}}{\partial x^\alpha}+\Gamma^\alpha_{\mu\gamma}\Gamma^\gamma_{\alpha\nu}-\Gamma^\alpha_{\mu\nu}\frac{\partial\ln\sqrt{-g}}{\partial x^\alpha}\,.$$

In a geodesic system of co-ordinates we can always make the $\Gamma^\alpha_{\beta\gamma}$ vanish at a point so that at this point (or along a geodesic) the Ricci tensor reduces to

$$(5.10)\qquad R_{\mu\nu}=\frac{\partial^2\ln\sqrt{-g}}{\partial x^\mu\,\partial x^\nu}-\frac{\partial\Gamma^\alpha_{\mu\nu}}{\partial x^\alpha}\,.$$

Another useful form is to introduce « proper co-ordinates », for which by definition $\sqrt{-g}=1$, one then has

$$(5.11)\qquad R_{\mu\nu}=-\frac{\partial\Gamma^\alpha_{\mu\nu}}{\partial x^\alpha}+\Gamma^\alpha_{\mu\gamma}\Gamma^\gamma_{\alpha\nu}\,.$$

It is perhaps interesting, historically, to remark that all of Einstein's 1915 papers were based on using « proper co-ordinates », even through 1918 proper co-ordinates were used extensively. One of the simplifying aspects of proper co-ordinates is that many expressions are the same as those in special relativity, for example, the Maxwell divergence equations $\partial_\nu(\sqrt{-g}\,F^{\mu\nu})=\sqrt{-g}\,j^\mu$ are the same as in special relativity; also $\sqrt{-g}\,\mathrm{d}^4x\to\mathrm{d}^4x$; etc. Thus a great simplification is achieved. On the other hand it turns out that when it comes to solving the field equations, it is usually simpler to introduce co-ordinates that are not proper, but related to proper co-ordinates by an Euclidean map, *i.e.*, in the same way that polar co-ordinates are related to rectangular co-ordinates. For some solutions even this device proves inconvenient. Some authors (notably Fock) prefer « harmonic co-ordinates » for which $\partial_\nu(\sqrt{-g}\,g^{\mu\nu})=0$. Note that in contrast, proper co-ordinates satisfy $\partial_\nu(\sqrt{-g}\,\delta^\nu_\mu)=0$ and are therefore less restrictive.

Needless to say one would be led to logical inconsistencies if one tried to argue for a particular choice of co-ordinate conditions to be applied independently of the symmetry of the problem. For example, consider the classical

490

equation in flat space $\nabla\cdot\boldsymbol{V}=0$, where $\boldsymbol{V}$ is some vector. In covariant notation this can be written $(\partial/\partial x_i)(\sqrt{\gamma}\,\gamma^{ij}V_j)=0$, where γ_{ij} is defined of course by $\mathrm{d}\sigma^2=\gamma_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$. The « harmonic co-ordinate conditions » for this problem would demand $\partial_i(\sqrt{\gamma}\,\gamma^{ij})=0$. However if we now suppose the problem has spherical symmetry, it is clearly appropriate to introduce spherical polar co-ordinates which do not satisfy either harmonic or proper co-ordinate conditions. Clearly the difficulty lies with the co-ordinate conditions rather than with the choice of co-ordinates.

5·2. – Einstein's field equations.

Let us now attempt to summarize the preceding ideas of general relativity in a series of postulates which will suggest in a more formal way how we should go about constructing the field equations. No attempt will be made in this summary to state the postulates in a mathematically rigorous or concise way.

Postulate 1. The laws of nature can be expressed in a generally covariant way, treating time as a co-ordinate along with the three spatial co-ordinates.

Postulate 2. Space and time form a four-dimensional manifold for which there is an element of distance (in analogy with Riemannian geometry)

$$\mathrm{d}s^2=g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \qquad (\mu,\nu=0,1,2,3)$$

with the interpretation given previously of $\mathrm{d}s^2$. In particular, we can always choose co-ordinates so that *locally*, $\mathrm{d}s^2=\mathrm{d}T^2-\mathrm{d}X^2-\mathrm{d}Y^2-\mathrm{d}Z^2$, that is, special relativity holds locally in a sufficiently small region of the space-time manifold just as in Riemannian geometry, with $g_{\mu\nu}$ positive definite, Euclidean geometry holds locally.

Postulate 3. The element of distance also determines the dynamics of a particle (generalizing from special relativity) according to the variational principle:

$$\delta\int m\,\mathrm{d}s=0\,,$$

together with $\mathrm{d}s=0$, for a light ray. Since m plays no role in this variational principle we have here a partial expression of the principle of equivalence. (Remark: It has been shown by later work that the equations of motion of particles follow from the field equations so that this postulate is not necessary.)

Postulate 4. *The principle of equivalence*: Just as in an accelerated co-ordinate system (*e.g.* the rotating co-ordinate system previously considered)

there are « inertial forces » arising as a consequence of the fact that the $g_{\mu\nu}$ are different from the $\eta_{\mu\nu}$ of special relativity, so too in the presence of permanent gravitational fields the $g_{\mu\nu}$ will differ from the $\eta_{\mu\nu}$.

Postulate 5. *Mach's principle*: As a consequence of the relativity of motion the inertial behavior of a body should be determined by the presence of other bodies in the universe. Since the dynamics of a body follows from Postulate 3, it follows that the metrical properties of the space-time manifold should be determined by the distribution of energy and momentum in the universe.

Postulate 6. *Relativistic correspondence principle*: In the non-relativistic limit, Newtonian mechanics and gravitational theory hold.

To construct the field equations using these postulates we proceed as follows: From Postulate 1, we know that the field equations will involve some tensors of various ranks and will be of the form

$$A^{\mu\nu\ldots}_{\alpha\beta\ldots} = B^{\mu\nu\ldots}_{\alpha\beta\ldots}\,.$$

We also know that since we are to determine the $g_{\mu\nu}$ from the equations (postulate 4) the equations should at least involve ten quantities corresponding to the ten $g_{\mu\nu}$. However since by a co-ordinate transformation we can always eliminate in general four of the $g_{\mu\nu}$ (more conveniently we can make $g_{00}=1$, $g_{0i}=0$), these equations should satisfy four identities so that they are underdetermined with respect to the $g_{\mu\nu}$ and permit the freedom of four arbitrary co-ordinate transformations. Using Postulate 5 we should expect to see the energy-momentum tensor $T_{\mu\nu}$ of matter play the role of being a source of the field. And finally from Postulate 6 we should expect that in the non-relativistic limit the field equations should reduce to Poisson's equation for the potential

$$\nabla^2 U = 4\pi G \frac{T_{00}}{c^2} = 4\pi G\varrho\,, \tag{5.12}$$

where ϱ is the matter density.

A system of equations having this property is the following

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = -\varkappa T_{\mu\nu}\,, \tag{5.13}$$

where $\varkappa \equiv 8\pi G/c^4$. We see that it satisfies four identities (from the previous section)

$$G^{\mu\nu}{}_{;\nu} = -\varkappa T^{\mu\nu}{}_{;\nu} = 0\,. \tag{5.14}$$

The fact that $T^{\mu\nu}{}_{;\nu}=0$ follows from the conservation law of energy and momentum in special relativity,

$$\frac{\partial T^{\mu}_{\nu}}{\partial x^{\mu}}=0\,. \tag{5.15}$$

This equation is not of course generally covariant but its covariant analog is

$$\sqrt{-g}\,T^{\mu}_{\nu;\mu}=\frac{\partial\sqrt{-g}\,T^{\mu}_{\nu}}{\partial x^{\mu}}-\frac{1}{2}\frac{\partial q_{\lambda\mu}}{\partial x^{\mu}}\,T^{\lambda\nu}\sqrt{-g}=0\,. \tag{5.16}$$

In the case of special relativity $\sqrt{-g}=1$, $\partial g_{\lambda\nu}/\partial x^{\mu}=0$, and this reduces to the above. To show that (5.13) reduces to Poisson's equation it is convenient to linearize the equations setting

$$g_{\mu\nu}=n_{\mu\nu}+h_{\mu\nu}\,, \tag{5.17}$$

with $h_{\mu\nu}$ regarded as small.

Proceeding in this way and adopting suitable co-ordinate conditions one can in fact show (BERGMANN, p. 182-185; MOLLER, p. 313-315) that

$$\nabla^2\frac{h_{00}}{2}=\frac{\varkappa\varrho}{2}=4\pi G\varrho\,. \tag{5.18}$$

For the case of a mass point ϱ is a δ-function and outside this point we have Laplace's equation $\nabla^2 h_{00}=0$. It follows from (5.18) that to lowest order the metric is given by (introducing suitable units)

$$\mathrm{d}s^2=\left(1-\frac{2GM}{rc^2}\right)c^2\,\mathrm{d}t\,z-\mathrm{d}r^2-r^2\sin^2\theta\,\mathrm{d}\varphi^2-r^2\,\mathrm{d}\theta^2\,, \tag{5.19}$$

as we obtained previously using the principle of equivalence.

However, we should like to emphasize that these considerations were based on using a « weak-field » approximation and a suitable choice of co-ordinates. The exact structure of the field equations indicates one can obtain T^0_0 directly from g^{rr} (no summation) via a *first order* linear differential equation in the co-ordinate system used below, a Laplacian operator on g_{00} showing up in the equation for the *transverse stress*. Thus some caution is necessary in drawing inference from the weak-field perturbation method.

5·3. – The Schwarzschild field.

Let us now obtain the solution to the field equations for a point mass, that is we look for solutions to

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=0\,. \tag{5.20}$$

This being the generalization of Laplace's equation in general relativity. Note that since $g^{\mu\nu}R_{\mu\nu}=R$ the above equations also imply $R=0$, $R_{\mu\nu}=0$. It should be pointed out that although $R_{\mu\nu}=0$, it does not follow $R^{\alpha}{}_{\mu\lambda\nu}=0$, corresponding to flat space. The equations $R_{\mu\nu}=0$ represent ten equations, whereas the equations $R^{\alpha}{}_{\mu\lambda\nu}=0$ represent twenty equations. Only in the case of 2-dimensional space-time or 3-dimensional space-time would the equations $R_{\mu\nu}=0$ imply space-time is flat. Thus the field equations make no sense unless space-time has at least $3+1$ dimensions (*).

To simplify the solution to the field equations we demand that the $g_{\mu\nu}$ be independent of time, and the line element spherically symmetrical. One can show that under these circumstances it can be brought into the form

$$ds^2 = e^{\nu}\,dt^2 - e^{\lambda}\,dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\varphi^2\,, \tag{5.21}$$

where $\nu=\nu(r)$, $\lambda=\lambda(r)$. Such a line element is described as a static line element with spherical symmetry. Needless to say any transformation of the form $r=f(r^*)$, $\theta^*=\theta$ do not destroy the spherical symmetry. To solve the equations we now compute the Christoffel symbols corresponding to this line element and substitute in the field equations. It is most convenient to use the field equations in mixed form $R^{\mu}_{\nu}-\frac{1}{2}\delta^{\mu}_{\nu}R=0$. We have, in general, after computing the Christoffel symbols and substituting in the field equations (**)

$$\left\{\begin{aligned} &-\exp[-\lambda]\left(\frac{\nu'}{r}+\frac{1}{r^2}\right)+\frac{1}{r^2}=\varkappa T^1_1\,,\\ &-\exp[-\lambda]\left(\frac{\nu''}{2}-\frac{\lambda'\nu'}{4}+\frac{\nu'^2}{4}+\frac{\nu'-\lambda'}{2r}\right)=\varkappa T^2_2=\varkappa T^3_3\,,\\ &\exp[-\lambda]\left(\frac{\lambda'}{r}-\frac{1}{r^2}\right)+\frac{1}{r^2}=\varkappa T^0_0\,.\end{aligned}\right. \tag{5.22}$$

and we obtain the solution to the desired case upon setting $T^1_1=T^2_2=T^3_3=$ $=T^0_0=0$. Subtracting the third from the first equation we have

$$-\exp[-\lambda]\left(\frac{\nu'}{r}+\frac{\lambda'}{r}\right)=0\,, \tag{5.23}$$

and $\nu=-\lambda+$ a constant. The third equation can be solved directly for λ;

(*) Whether the field equations and equations of motion derived from the geodesic postulate make sense for $\delta^{\mu}_{\mu}>4$ is of course an important question, since if they did not, we should have a basis in general relativity for answering the frequently asked question, « why is $\delta^{\mu}_{\mu}=4$? ».

(**) As done, for example, in Tolman's book.

we have, putting $v = e^{-\lambda}$, $v' = -e^{-\lambda}\lambda'$ and substituting

$$v' + \frac{v}{r} = \frac{1}{r}\,, \tag{5.24}$$

the solution to which is $v = 1 + (A/r)$. Now since we have $e^{\nu} = ke^{-\lambda} = k(1 + (A/r))$ and since we know from the principle of equivalence $\mathrm{d}s^2 = e^{\nu}\,\mathrm{d}t^2 \approx (1 - (2GM/rc^2))\,\mathrm{d}t^2$, it follows we should choose $k = 1$, $A = -2GM/rc^2$ and $\mathrm{d}s^2$ becomes

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right)c^2\,\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{(1 - 2GM/rc^2)} - r^2\,\mathrm{d}\theta^2 - r^2\sin^2\theta\,\mathrm{d}\varphi^2\,. \tag{5.25}$$

The solution with $g_{\mu\nu}$ as above (apart from a co-ordinate transformation is called « Schwarzschild's exterior solution ». He also obtained a solution for the region inside a spherical distribution of matter, called the « interior solution », which we shall not discuss here.

The above discussion represents the customary approach to solving the field equations; it is clearly very formal and does not give much physical insight into the problem. Since these notes were first written we have been led to proceed in a slightly different manner (*Phys. Rev. Lett.*, **6**, 147 (1961)). The first thing to notice is that if we replace $-e^{\nu}$ by g^{11} (*i.e.* g^{rr}), the equation for T^0_0 is linear and first order in g^{11}! We note also that since the equations indicate together with the boundary conditions that $g_{00}g_{11} = -1$, we may eliminate these two quantities in favor of the scalar potential defined by $-g^{11} = g_{00} = 1 + 2U$. We then find outside matter, or more generally if $T^0_0 = T^1_1$, the two *dependent* equations (*)

$$\frac{2}{r^2}\left[\frac{\mathrm{d}}{\mathrm{d}r}(rU)\right] = -\varkappa T^0_0\,, \qquad (= -\varkappa T^1_1)\,.$$

$$\nabla^2 U = -\varkappa T^2_2\,, \qquad (= -\varkappa T^3_3)\,,$$

where ∇^2 is the flat-space radial Laplacian. Thus for the only non-trivial solution that has ever been checked experimentally, the Schwarzschild field, one of the field equations of general relativity is identical with that of the Newtonian potential theory $\nabla^2 U = 0$, and the other equation an integral of it—since $(\mathrm{d}/\mathrm{d}r)(r^2G^0_0) - 2rG^2_2 = 0$. We plan to discuss this curious result in greater detail in a subsequent publication. Needless to say, this simple linear structure is not obtained if we adopt, *e.g.* isotropic radial co-ordinates, that is, co-ordinates

(*) We use the labels 1, 2, 3 to designate the co-ordinates r, θ, φ.

for which the spatial line element is

$$d\sigma^2 = f(\bar{r})[d\bar{r}^2 + \bar{r}^2(d\theta^2 + \sin^2\theta\, d\varphi^2)]\,,$$

or if we use proper radial distance, *i.e.* $\varrho = \int\sqrt{-g_{11}}\,dr$.

The co-ordinate « r » that we are using is a proper transverse measure. It is proportional to the perimeter of a spherical surface surrounding the Schwarzschild singularity when measured by infinitesimal rods tangent to the surface along a great circle. A rod oriented in this fashion preserves its length unchanged relative to the system of co-ordinates, and *a fortiori* relative to an observer at infinity, in contrast with a rod oriented radially. Thus in introducing « r » we have introduced an invariant transverse measure of radial distance: a spherical shell is said to be closer or further from the singularity relative to another spherical shell of perimeter $2\pi r_0$ depending on whether $r < r_0$ or $r > r_0$. From our work in the previous section, it is clear $r = K^{-\frac{1}{2}}$, where K is the Gaussian curvature of the spherical surface. Since K is the product of the principal curvatures (which are the same for a sphere) we have therefore labelled the spheres according to their radius of curvature.

5·4. – Orbital equations in a Schwarzschild field.

Having obtained the expression for the line element of the Schwarzschild solution, we are now in a position to solve the equations of motion obtained from the variational principle $\delta\int ds = 0$. The general equations are

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\lambda\nu}\frac{dx^\lambda}{ds}\frac{dx^\nu}{ds} = 0\,. \tag{5.26}$$

We may now compute the Christoffel symbols corresponding to the line element and hence obtain explicit expressions in the above. However it is possible to solve this problem *without ever computing Christoffel symbols* using the first integrals of the equations of motion.

First let us assume the motion takes place in the plane $\theta = \pi/2$ so that the line element is simply (dropping the c)

$$ds^2 = \left(1 - \frac{2GM}{r}\right)dt^2 - g_{00}^{-1}\,dr^2 - r^2\,d\varphi^2\,. \tag{5.27}$$

Now in the variational principle, since ds is invariant under $\varphi \to \varphi + \text{const}$, $t \to t + {}$ $+ \text{const}$, it follows that there will be two conserved quantities, corresponding to energy and angular momentum.

Alternatively, in the equations below, we have $\partial L/\partial t = \partial L/\partial\varphi = 0$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{\partial L}{\partial \dot t} - \frac{\partial L}{\partial t} = 0\,, \qquad \frac{\mathrm{d}}{\mathrm{d}s}\frac{\partial L}{\partial \dot\varphi} - \frac{\partial L}{\partial \varphi} = 0\,. \tag{5.28}$$

Hence $\partial L/\partial \dot t = K_0$, $\partial L/\partial\dot\varphi = K_\varphi$. Performing the differentiations we have

$$\begin{cases} \left(1 - \dfrac{2GM}{r}\right)\dfrac{\mathrm{d}t}{\mathrm{d}s} = K_0\,, \\ r^2\dfrac{\mathrm{d}\varphi}{\mathrm{d}s} = K_\varphi\,. \end{cases} \tag{5.29}$$

The first integral is effectively the energy conservation law, and the second, Kepler's law. To see the former, expand $\mathrm{d}s$ then we obtain

$$\mathrm{d}s = \left(1 - \frac{2GM}{r}\right)^{\frac{1}{2}}\mathrm{d}t\left[1 - \frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 - \frac{1}{2}r^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 + \dots\right], \tag{5.30}$$

and substituting in (5.29) and again keeping the lowest order terms we have

$$1 + \frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + \frac{1}{2}r^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 - \frac{GM}{r} + \dots = K_0\,. \tag{5.31}$$

The first term is nothing but the rest energy of the particle mc^2 if we multiply through by m and introduce suitable units. Thus $K_0 - 1$ plays the role of the term E/m we introduced in an earlier discussion. Let us now rewrite the expression for the line element in the following form

$$1 = \left(1 - \frac{2GM}{r}\right)\left(\frac{\mathrm{d}t}{\mathrm{d}s}\right)^2 - g_{00}^{-1}\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 - r^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}s}\right)^2, \tag{5.32}$$

and simplify using (5.29). Then we have using also $\mathrm{d}r/\mathrm{d}s = (\mathrm{d}r/\mathrm{d}\varphi)(\mathrm{d}\varphi/\mathrm{d}s)$

$$1 = K_0^2 g_{00}^{-1} - g_{00}^{-1}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\frac{K_\varphi^2}{r^4} - \frac{K_\varphi^2}{r^2}\,, \tag{5.33}$$

which may also be written

$$\left(1 - \frac{2GM}{r}\right) = K_0^2 - \left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\frac{K_\varphi^2}{r^4} - \frac{K_\varphi^2}{r^2}\left(1 - \frac{2GM}{r}\right). \tag{5.34}$$

Now introduce $u = r^{-1}$ and the above becomes

$$1 - 2GMu = K_0^2 - \left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 K_\varphi^2 - K_\varphi^2 u^2 + K_\varphi^2\cdot 2GMu^3\,. \tag{5.35}$$

Taking the derivative we obtain

$$\frac{d^2u}{d\varphi^2} + u = \frac{GM}{K_\varphi^2} + 3GMu^2 . \tag{5.36}$$

The non-linear term on the right is the correction to the usual Newtonian equations of the orbit represented by the first three terms. Thus setting $K_\varphi = K'_\varphi/c$, $G \to G/c^2$ the equation can be written as

$$\frac{d^2u}{d\varphi^2} + u = \frac{GM}{K'^2_\varphi} + \frac{3GM}{c^2} u^2 , \tag{5.37}$$

and hence in the limit $c \to \infty$ the Newtonian equation results in accordance with the relativistic correspondence principle.

Let us now look for a perturbative solution to (5.37) treating the last term as small. We may imagine the orbit is approximately circular and hence to a first approximation $u = \text{const} = GM/K'^2_\varphi$. Hence we look for a solution of the form

$$u = \frac{GM}{K'^2_\varphi} + u_1 , \tag{5.38}$$

and (5.37) becomes, neglecting higher order terms,

$$\frac{d^2u_1}{d\varphi^2} + \left(1 - \frac{6G^2M^2}{c^2K'^2_\varphi}\right) u_1 = 0 . \tag{5.39}$$

Since in the Newtonian case the solution would be

$$u = GM/K'^2_\varphi + (\text{constant}) \cos\varphi \tag{5.40}$$

corresponding to an ellipse, we see that the lowest order effect is simply to modify the Newtonian solution to a precessing ellipse, that is,

$$u = \frac{GM}{K'^2_\varphi} + (\text{constant}) \cos \sqrt{1 - \frac{6G^2M^2}{c^2K'^2_\varphi}}\, \varphi . \tag{5.41}$$

After φ has gone through 2π it must still go through an additional amount for u to return to its original value at $\varphi = 0$; we have, expanding the radical

$$\left\{ \begin{aligned} &\left[1 - \frac{3G^2M^2}{c^2K'^2_\varphi} + \ldots\right] (2\pi + \Delta) = 2\pi , \\ &\Delta = 2\pi \cdot \frac{3G^2M^2}{c^2K'^2_\varphi} \text{ rad} . \end{aligned} \right. \tag{5.42}$$

408

TABLE II. – *Theoretical values of the advance of the perihelia per century (e is the eccentricity of the ellipse).*

Planet	$\overline{\omega}'$	$e\overline{\omega}'$
Mercury	43″.03	8″.847
Venus	8 .63	0 .059
Earth	3 .84	0 .064
Mars	1 .35	0 .126
Jupiter	0 .06	0 .003

TABLE III. – *Contributions to the motion of the perihelia of Mercury and the earth.*

Cause		Motion of perihelion	
	m^{-1} (*)	Mercury	Earth
Mercury	6 000 000 ± 1 000 000	0″.025 ± 0″.00	– 13″.75 ± 2″.3
Venus	408 000 ± 1 000	277 .856 ± 0 .68	345 .49 ± 0 .8
Earth	329 390 ± 300	90 .038 ± 0 .08	—
Mars	3 088 000 ± 3 000	2 .536 ± 0 .00	97 .69 ± 0 .1
Jupiter	1 047.39 ± 0.03	153 .584 ± 0 .00	696. 85 ± 0 .0
Saturn	3 499 ± 4	7 .302 ± 0 .01	18 .74 ± 0 .0
Uranus	22 800 ± 300	0 .141 ± 0 .00	0 .57 ± 0 .0
Neptune	19 500 ± 300	0 .042 ± 0 .00	0 .18 ± 0 .0
Solar oblateness	—	0 .010 ± 0 .02	0 .00 ± 0 .0
Moon	—	—	7 .68 ± 0 .0
General precession (Julian century, 1850)	—	5025 .645 ± 0 .50	5025 .65 ± 0 .5
Sum	—	5 557 .18 ± 0 .85	6 179 .1 ± 2 .5
Observed motion	—	5 599 .74 ± 0 .41	6 183 .7 ± 1 .1
Difference	—	42 .56 ± 0 .94	4 .6 ± 2 .7
Relativity effect	—	43 .03 ± 0 .03	3 .8 ± 0 .0

(*) (m^{-1} = reciprocal of the adopted masses, the sun's mass taken as unity.

Note that with improved astronomical observations and calculations (say by a factor of ten), the precession of Venus, Earth and Mars could also be checked with the theory. (Our notes.)

The quantity Δ represents the angular distance, therefore, between two successive perihelions. We see that in the limit $c^{-1} \to 0$, $\Delta \to 0$, in accordance with our relatvistic correspondence principle. Inserting the values of the various quantities above, we find for Mercury $\Delta \approx 43''$ per century. This esti-

mated precession is in good agreement with that found astronomically after the precession due to planetary perturbations has been taken into account. It was first calculated by LEVERRIER in the 19th century, who, on the basis of the discrepancy postulated the existence of another planet to account for the discrepancy. No such planet was found, and this constituted a fundamental difficulty for Newtonian mechanics. The experimental situation is summarized in Tables II and III (from G. M. CLEMENCE: *Rev. of Mod. Phys.*, **19** 361 (1947)).

Various attempts were made to modify the Newtonian theory, but with little success, since it is clearly unsatisfactory to change a theory as fundamental as the Newtonian theory merely to predict *one effect*. With the advent of special relativity, it was appreciated that simply because of the increase of inertial mass with velocity there would be a precession. However, as we shall now show, using a scalar theory of gravitation, the precession is only $\frac{1}{6}$ of the observed value and of the wrong sign.

In accordance with special relativity and the hypothesis of a scalar gravitational field we postulate a variational principle of the form:

$$\delta\int m\,\mathrm{d}\tau + mU\,\mathrm{d}\tau = 0\,, \tag{5.43}$$

where $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$ and U is the gravitational potential $U = -GM/r$ for a « fixed source » of mass M. We have again the two conservation laws (upon introducing polar co-ordinates for $\mathrm{d}\tau^2$ as above for $\mathrm{d}s^2$)

$$\left\{\begin{aligned} \frac{\partial L}{\partial \dot{t}} &= (1+U)\frac{\mathrm{d}t}{\mathrm{d}\tau} = K_0\,,\\ \frac{\partial L}{\partial \dot{\varphi}} &= (1+U)r^2\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = K_\varphi\,. \end{aligned}\right. \tag{5.44}$$

Substituting in the expression for $\mathrm{d}\tau^2$

$$1 = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 - \left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - r^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\tau}\right)^2, \tag{5.45}$$

using (5.44), we find

$$(1+U)^2 = K_0^2 - \left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2\frac{K_\varphi^2}{r^4} - \frac{K_\varphi^2}{r^2}\,. \tag{5.46}$$

Now introducing $u = r^{-1}$ and expanding the left-hand side we obtain

$$1 - 2GMu + G^2M^2u^2 = K_0^2 - \left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 K_\varphi^2 - K_\varphi^2u^2\,. \tag{5.47}$$

Differentiating and once again introducing ordinary units we have

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\varphi^2} + \left(1 + \frac{G^2 M^2}{c^2 K_\varphi'^2}\right) u = \frac{GM}{K_\varphi'^2}\,. \tag{5.48}$$

We see the left hand side is the same as our previous eq. (5.39), except that the precession term is $\frac{1}{6}$ of the correct value and is of the wrong sign.

This result constituted a serious difficulty, as we have indicated previously, for any attempts to build a theory of gravitation on a special relativistic basis. (*) This failure of an approach based on special relativity was to be expected, however, on more fundamental grounds.

5·5. – The gravitational deflection of light.

In Chapter III, using the principle of equivalence we calculated the deflection of light by a gravitating body treating it as a Newtonian body of mass $h\nu/c^2$. We now wish to perform the calculation using the correct expression for the line element, by directly solving the equations:

$$\begin{cases} \delta\displaystyle\int \mathrm{d}s = 0\,, \\ \mathrm{d}s = 0\,. \end{cases} \tag{5.49}$$

If we examine the constants of the motion (5.29) we see that they both become infinite in the limit $\mathrm{d}s \to 0$. Turning to the orbital eq. (5.37), we see that it only involves K_φ' and hence letting $K_\varphi' \to \infty$, the equation becomes

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\varphi^2} + u = \frac{3GM}{c^2}\, u^2\,. \tag{5.50}$$

Now in the limit $c^{-1} \to 0$, this is the equation of a straight line, similarly in the limit $G \to 0$. Thus the effect of the gravitational field is to destroy the Euclidean straight-line character of the propagation of light rays in free space. The null geodesics $\mathrm{d}s = 0$ being those which satisfy the non-linear eq. (5.50).

To solve this equation to desired accuracy we set as a first approximation, $u_0 = R^{-1}\cos\varphi$, where R is the distance of closest approach to the gravitating body. We have therefore $u = u_0 + u_1 + \ldots$, substituting this in (5.50) we find

$$\frac{\mathrm{d}^2 u_1}{\mathrm{d}\varphi^2} + u_1 = \frac{3GM}{c^2 R^2}\cos^2\varphi\,. \tag{5.51}$$

(*) For a discussion of more complicated attempts see *e.g.*, S. N. Gupta: *Rev. Mod. Phys.*, **29**, 334 (1957).

Trying a solution of the form $u_1 = A \sin^2\varphi + B$ we obtain $A = G^2M^2/c^2R^2$, $B = A$, and u is therefore given by

$$u = \frac{1}{R}\cos\varphi + \frac{GM}{c^2R^2}(\sin^2\varphi + 1)\,. \tag{5.52}$$

Along the asymptotes we have as before

$$\left\{\begin{aligned} 0 &= \frac{1}{R}\cos\varphi_1 + \frac{GM}{c^2R^2}(\sin^2\varphi_1 + 1)\,,\\ 0 &= \frac{1}{R}\cos\varphi_2 + \frac{GM}{c^2R^2}(\sin^2\varphi_2 + 1)\,, \end{aligned}\right. \tag{5.53}$$

Now since $\varphi_1 = (\pi/2) + \Delta$, where Δ is small, substituting we find $\Delta = 2GM/c^2R^2$ and the total deflection is

$$\alpha = 2\Delta = \frac{4GM}{c^2R^2}\,, \tag{5.54}$$

or twice the total deflection found previously. It is interesting to note also that the velocity of the light does *not* increase as it approaches the sun as for a Newtonian particle, but diminishes. This follows from taking the ratio of the two expressions in (5.29); we have

$$r^2\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{K_\varphi}{K_0}\left(1 - \frac{2GM}{r}\right), \tag{5.55}$$

and for $r = R$, this becomes

$$R^2\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{K_\varphi}{K_0}\left(1 - \frac{2GM}{R}\right). \tag{5.56}$$

But on the other hand at $r = R$ from the line element, since $\mathrm{d}r/\mathrm{d}t = 0$ we have

$$R^2\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 = 1 - \frac{2GM}{R}\,. \tag{5.57}$$

Hence eliminating $\mathrm{d}\varphi/\mathrm{d}t$ and solving for K_φ/K_0 and then restoring units we find

$$r^2\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{Rc}{(1 - (2GM/Rc^2))^{\frac{1}{2}}}\left(1 - \frac{2GM}{rc^2}\right). \tag{5.58}$$

Thus at perihelion the areal velocity is not Rc in terms of a clock at infinity but

$$R^2\frac{\mathrm{d}\varphi}{\mathrm{d}t} = Rc\left(1 - \frac{2GM}{rc^2}\right)^{\frac{1}{2}}. \tag{5.59}$$

412

Although the local velocity in terms of a clock at rest at R is of course c *i.e.*,

$$\frac{R}{(1-(2GM/Rc^2))^{\frac{1}{2}}}\frac{\mathrm{d}\varphi}{\mathrm{d}t}=c\,. \tag{5.60}$$

The experimental situation on the deflection of light is summarized in the following table taken from R. J. TRUMPLER (op. cit., p. 108).

TABLE IV. – *Observation of light deflection.*

Eclipse	Station	Focal length (feet)	No. plates	No. stars	Meth-od	Light Deflection at Sun's Limb	p.e.	Observers
1919 May 29	Sobral	19	7	7	relat.	1.98	±.12	DYSON
	Sobral	11	16	6 ÷ 12	relat.	(0.86)	±.1	DAVIDSON
	Principe	18	2	5	relat.	1.61	±.3	EDDINGTON
1922 Sept. 21	Wallal	15	4	62 ÷ 85	relat.	1.72	±.11	CAMPBELL
		5	6	134 ÷ 143	relat.	1.82	±.15	TRUMPLER
	Wallal	10	2	18	relat	1.74	±.3	CHANT, YOUNG
	Cordillo Downs	5	2	14	relat.	1.77	±.3	DAVIDSON, DODWELL
1929 May 9	Takengon	28	4	17 ÷ 18	absol.	(2.24)		FREUNDLICH
					relat.	1.75	±.13	v. KLÜBER
								v. BRUNN
1947 May 20	Bocaiuva	20	1	51	relat.	2.01	±.18	v. BIESBROEK
1952 Febr. 25	Khartoum	20	2	9 ÷ 11	absol.	1.70	±.07	v. BIESBROEK

Prediction by Relativity Theory 1.75.

These two results (precession of the perihelion of Mercury, together with the deflection of light) constitute the other checks of general relativity beyond the change of a clock's rate in a gravitational field discussed earlier, and which follow immediately from our expression for ds.

One customarily speaks of these effects as the « three checks » of general relativity. It should be noted that the precession of the perihelion of Mercury is a very severe check on the theory since it is of order G^2 in the calculations, the other effects being of order G. It is to be hoped that the development of experimental techniques in the coming years will make it possible to verify this precession effect for other planets and satellites, and also open new ways to measure effects of order G^2.

5·6. – Some orbital formulae common to Newtonian mechanics and general relativity.

Another « check » of general relativity not often discussed appears to be the remarkable similarity of certain expressions obtained in the theory with those of Newtonian mechanics. We have already commented previously on this in connection with the field equations. Two striking instances occur in the equations of motion in addition to the Keplerian areal law mentioned previously:

1) For a particle falling radially, the acceleration is given by

$$\frac{\mathrm{d}^2 r}{\mathrm{d}s^2} = -\frac{GM}{r^2}\,. \tag{5.61}$$

where of course $\mathrm{d}s$ measures the interval of time read by a clock falling with the particle. We see that if one measures time by such a freely-falling clock, and « distance » by the invariant transverse measure (radius of curvature of the spherical shells through which the particle-clock is falling), then the formula is identical with Newtonian mechanics. Moreover a Newtonian could not object to this interpretation of the law since these constitute alternative and admissible ways to make the measurement.

2) Let a particle be undergoing circular motion, for which $\mathrm{d}^2 r/\mathrm{d}s^2 = 0$, then from the geodesic equations of motion we have

$$\Gamma^r_{00}\,\mathrm{d}t^2 + \Gamma^r_{\varphi\varphi}\,\mathrm{d}\varphi^2 = 0 \tag{5.62}$$

or

$$\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 = \frac{1}{2r}\,g_{00,r} = \frac{GM}{r^3}\,. \tag{5.63}$$

in other words Kepler's third law holds exactly. This latter result would appear at first glance simply a coincidence due to our choice of co-ordinates (see *e.g.* EDDINGTON, p. 89). However a more careful inspection reveals that this expression is also identical with $R^{\varphi 0}{}_{\varphi 0}$, which in turn is one of the diagonal components of the Riemann curvature tensor expressed in canonical form for the Schwarzschild field (see the discussion of the Petrov classification of the Riemann tensor given in F. A. E. PIRANI: *Phys. Rev.*, **105**, 1089 (1957)). The relationship therefore has fundamental significance. Similarly, upon computing

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{dr^2}{\mathrm{d}s^2}\right) = \frac{2\,GM}{r^3}\,.$$

we find this is the same as $R^{r0}{}_{r0}$.

It would therefore appear from the above that a geometer could show that in formulating Newtonian mechanics, the Newtonian was using more as-

414

sumptions in his space-time geometry than were necessary, and if we strike away some of these asumptions (*e.g.* the rate of a clock in motion is the same as a clock at rest; the co-ordinate « r » represents proper radial distance as well as radius of curvature), the resulting assumptions (or conclusions based on observations) agree with those of general relativity. In other words, there is a residual body of statements common to both theories, just as is the case for Euclidean and non-Euclidean geometry. To the extent that a particular result depends only on what the two theories have in common, the results will be the same. The above two results appear to be of this character.

Needless to say, such a « check » of general relativity is of an aesthetic or logico-geometrical character rather than a direct empirical one; but in view of the inherent methodological difficulties of establishing any theory by empirical means alone and particularly when the predictions are so few in terms of what is experimentally possible (in contrast with quantum mechanics), it appears highly desirable to have supplementary means for criticizing and evaluating the many (not very successful) alternative theories to general relativity that have been proposed, as well to sharpen our understanding of the general theory itself.

However, the severest « check » of the theory will be whether or not it proves useful in helping us to arrive at a more unified and experimentally verifiable picture of microphysics and cosmology—to which we now turn.

CHAPTER VI.

MACH'S PRINCIPLE AND COSMOLOGY

6·1. – The clock paradox in the presence of a permanent gravitational field.

In the previous chapter we have discussed the « classical » three empirical checks of general relativity together with checks of a more mathematical character. However if these were the only kinds of checks of the theory it would be unsatisfactory: some checks of a cosmological character are needed.

The reason for this is the following: the conceptual basis of relativity lies in the assumption that motion is an entirely relative concept, and hence it is only meaningful to speak of the motion of a body (*) provided there are other

(*) The concept of a « body » is of course vague; but we usually mean something we can observe either directly or indirectly by physical means and which participates in the Laws of Nature in terms of energy balance, etc. Usually we mean something composed of atoms or the « elementary particles », but « fields » also may be subsumed in this definition.

bodies relative to which the motion may be referred. If we now consider, *e.g.* the famous clock paradox (*), in special relativity, we encounter a difficulty with respect to the relativity of motion unless there are other bodies in the universe besides the two clocks involved in the problem. That is to say, the asymmetry observed in the readings of the clocks is to be attributed to the fact that one of the clocks was accelerated relative to the other bodies in the universe.

One might at first ask: Why relative to the other bodies in the universe, why not consider acceleration as absolute? The answer to the latter question is twofold: *a*) if we adopt this solution we are giving up relativity, and *b*) this solution, as we shall now show, is untenable from the standpoint of gravitation and the principle of equivalence.

The proof runs as follows:

First we remind the reader that according to the principle of equivalence an observer in uniform acceleration cannot decide locally whether:

a) he is actually at rest on the surface of a gravitating body, or

b) in « free space » accelerating relative to the distant stars and in the limit as the acceleration vanishes (as measured locally by an accelerometer) he cannot determine whether

a) he is freely falling in a gravitational field, and hence accelerating relative to the distant stars, or

b) in « free space » translating uniformly relative to the distant stars.

We now use these ideas to construct a situation in which we get the « reverse » of the usual clock paradox.

Consider as in the diagram, a gravitating body

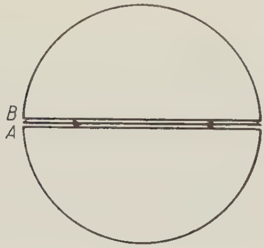


Fig. 5. – Spherical body with hole drilled through it. Clock B falls through the hole to the antipodal point and returns to clock A at rest on the surface where they are compared.

(*) Briefly stated the paradox is the following: we have two similar clocks A and B at rest relative to one another so that $\Delta t_A = \Delta t_B$ (*i.e.* a time interval of clock A = time interval of clock B). Let one clock now be moved or accelerated away from the other (in such a way that an accelerometer attached to one clock reads zero, while that attached to the other reads different from zero) and in such a way that it finally returns to the other clock. Then in general the clocks will not read the same, and in special relativity, the clock whose accelerometer reads zero throughout the experiment will read a greater elapsed time than the other clock.

with a hole drilled through the center from one antipode to the other. And let there be two clocks A and B with accelerometers attached. Now let clock A remain at rest at the surface of the hole and let B fall through the hole to the other antipode and then back to A where the clocks are compared. One can easily show by direct calculation (*) (this will be done by us in a paper to appear later this year) that the clock B will read less than the clock A upon its return. *Yet*, it was clock A, whose accelerometer read different from zero during the motion. From the standpoint of absolute acceleration (*i.e.* in terms of the reading of an accelerometer) the result is incredible, since from this viewpoint one would argue that clock B was the clock that constituted an inertial frame, and clock A, the one that did not, since it appeared to be accelerated as determined by its accelerometer readings. One would therefore conclude that clock A should read less than clock B which is incorrect.

On the other hand, if we compute the time difference between clock A and clock B for the situation where clock B instead of falling through the hole, is hurled outward from the mass, and with a sufficiently small velocity, so that it returns, then we find that the elapsed time read by clock A is *less than* the elapsed time indicated by B (as one would expect from a crude application of special relativity) (**). (In making the comparison we synchronize B with A, as B passes A, and *after* B has acquired initial acceleration, so that B is in free fall throughout the measurement).

By these considerations we are led to conclude that we cannot in general predict which of two clocks will read greater or less elapsed time after a relative round trip voyage. The relative readings of the two clocks will not be determined by their local apparent acceleration but by their integrated state of motion relative to the other bodies in the universe acting as sources of the gravitational field or metric.

In the above example, the « other bodies of the universe », were replaced by a large central field, presumably in the case of special relativity these other bodies, are the so-called « fixed-stars » whose net effect locally is to produce the special-relativity metric or the basic inertial frane, when local perturbations

(*) Using Schwarzschild's interior solution (see MØLLER, p. 330),

$$ds^2 = \left\{A - B\sqrt{1 - r^2/R^2}\right\}^2 c^2\, dt^2 - \frac{dr^2}{1 - r^2/R^2} - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)\,,$$

$$A = \tfrac{3}{2}\sqrt{1 - r_1^2/R^2}\,, \qquad B = \tfrac{1}{2}\,, \qquad R^2 = 3/\varkappa\varrho_0 c^2\,, \qquad r_1 = \text{radius of sphere}\,.$$

(**) The above results may be calculated to a sufficient approximation using the « flat earth » line element with $ds^2 = (1 + 2az)\, dt^2 - g_{00}^{-1}\, dz^2 - dx^2 - dy^2$, $a = GM/r_0^2$. In this co-ordinate system, for radial motion, the radial acceleration satisfies $d^2z/ds^2 = -a$, so that calculations are very simple. See also W. COCKRAN: *Vistas in Astronomy*, (Edited by A. BEER) London, 1960, p. 78.

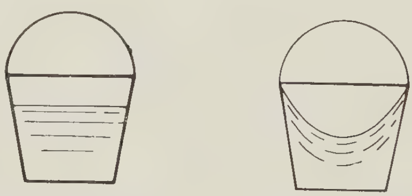
417

are neglected. In this case a well-defined asymmetry exists between the two clocks in the clock-paradox during the round trip and there is no contradiction. However this « explanation » makes sense only if the other bodies in the universe do in fact give rise to the Minkowski metric (to a good approximation) although, as we shall see, this assumption is not free of objection. Nevertheless, the assumption that the basic inertial frame is determined by some over-all effect of mass receives support from the fact that strong local concentrations of matter can seriously distort the inertial frame to the extent that uniform motion in a straight line becomes *e.g.* uniform circular motion (motion in a Schwarzschild field), or oscillatory motion on a straight line (such as the clock oscillating back and forth for the interior solution).

6·2. – Mach's principle.

The assumption that the distant stars define the basic inertial frame and that inertial effects therefore arise as a consequence of accelerations relative to the distant stars is called « Mach's principle ». MACH of course arrived at this idea, before the formulation of relativity theory, by reconsidering the famous argument given by NEWTON that rotation must be considered absolute.

NEWTON pointed out that if we consider some water in a rotating pail,



Pail in rotation. Water at rest.

Pail at rest. Water in rotation.

Fig. 6.

before the pail has transferred angular momentum to the water, the surface is horizontal, while after the pail has come to rest, and the angular momentum transferred to the water, the surface assumes a concave form. Now clearly this flatness or concavity of the water's surface has nothing to do with the relative rotation of water and pail, since in both cases, they are in relative rotation.

The cause therefore of the difference must be that the state of rotation (or rest) is with respect to some third body (or bodies). For NEWTON this third body was absolute space, however, MACH argued otherwise: suppose we imagine the pail to become larger and larger, that is to say, more massive, and

in fact, of the same dimensions and mass as the celestial bodies themselves, would it still be true that the relative rotation would play no role as to whether the surface of the water assumed a flat or concave shape? In other words, may not the distant stars in some manner play the role of the « third body »?

In Newtonian mechanics the answer is of course, « no », since the dynamical laws are formulated with respect to an absolute frame (to within a uniform translation *i.e.* invariance with respect to the Galilean group), and the only influence of the distant stars is to change the gravitational potential in the neighborhood of the rotating pail. No such effects envisaged by MACH are to be expected nor could be understood.

On the other hand, when we turn to general relativity there is such a possibility because of the tensorial character of the field. Indeed, THIRRING (*Phys. Zeits.*, **19**, 33 (1918)) showed that if we have a rotating shell in general relativity, then the gravitational field produced inside the shell has a metric of the following form (using the linearized approximation to the field equations)

$$g_{00} = 1 - \frac{2GM}{Rc^2}\left(1 + \frac{5R^2\omega^2}{3c^2} + \frac{\omega^2}{6c^2}(x^2+y^2-2z^2)\right),$$

$$g_{10} = \frac{4}{3}\frac{GM}{Rc^2}\frac{\omega y}{c},$$

$$g_{20} = -\frac{4}{3}\frac{GM}{Rc^2}\frac{\omega x}{c},$$

$$g_{30} = 0,$$

together with corrections to g_{ij} which may neglect.

As a consequence of that this solution we see that from the terms in g_{00} there will be a centrifugal force $\propto (GM/c^2R)\omega^2\varrho$, $\varrho^2 = x^2+y^2$, as well as Coriolis forces, from the terms g_{10}, g_{20}. Thus an observer inside the shell, at least in the plane $z=0$, would experience analogous effects to these experienced by an observer on a rotating disc. In particular if we put a small sphere of fluid at the center of the sphere, it would assume an ellipsoidal shape, as if it were rotating relative to Newton's absolute space.

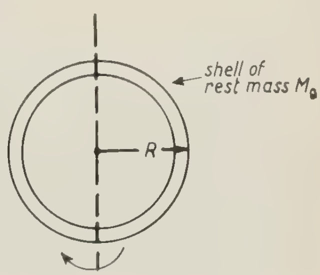


Fig. 7.

However, it is clear, this model of Thirring cannot fully explain Mach's principle, since we see that there are in g_{00} terms involving z which do not occur in a rotating system. Thus if we place a particle on the z-axis above the origin it will experience an inward directed acceleration. The reason for this has sometimes been attributed to the linear approximation used, but more

419

likely it is due to the neglect of the stress system in the sphere which maintains its spherical shape. Certainly however, the fact that the rotating sphere itself would tend to assume an elliptical shape means that we have not really got at the full explanation, since in the Thirring model there is no other matter in the universe except the sphere, *i.e.*, the energy-momentum tensor $T_{\mu\nu}$ only involves the sphere. Another difficulty is related to the fact that the Thirring model is intended for speeds $v \ll c$, whereas we have seen in Chapter II, that the kinematic or apparent velocities of the stars corresponds to $v \gg c$.

If we examine the structure of the field equations given previously, it is not difficult to see where the fundamental trouble lies. Thus in the absence of matter we have

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 0$$

for which $g_{\mu\nu} = \eta_{\mu\nu}$ (of special relativity) is a perfectly valid solution. However in this circumstance the equations of motion of a particle are obtained from $\delta \int (\eta_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu)^{\frac{1}{2}} = 0$ and if we transform to a rotating frame, we find the usual centrifugal and Coriolis forces, thus a particle can exhibit inertia even in the absence of distant masses and this is in clear contradiction with Mach's principle.

There is at present no accepted answer to this dilemma. We have attempted recently (*) a solution based on using the cosmological term $\Lambda g_{\mu\nu}$ which can be added to the field equations, since $\Lambda g_{\mu\nu;\nu} = 0$, however we prefer in these lectures to leave this point open as a matter of speculation for the reader and proceed to some of the other problems of cosmology.

6·3. – Olber's paradox and the finiteness of the universe.

Consider now the classical questions: « Is the universe finite or infinite? » « Are there a finite or infinite number of stars? » A simplified approach to these difficult questions is that given by OLBERS (1826) (**). Let us suppose: 1) we have an Euclidean universe with an infinite number of stars, 2) the stars are distributed with a uniform density ϱ, 3) every star has the same luminosity, 4) the stars have been radiating and absorbing light for an infinite time into the past. With these assumptions we are led to the conclusion that the sky must be infinitely bright in contradiction with observation.

(*) *Nuovo Cimento*, **20**, 1 (1961).

(**) For a more detailed account see BONDI, p. 19. Our presentation of this paradox differs somewhat from BONDI's one. In a recent review article, O. STRUVE (*Physics Today*, **14**, 54 (1961)) has called attention to an earlier statement of the paradox by L. DE CHESEAUX (1744).

To see this consider a series of spherical shells taken about a given point. Then the total number of stars in a given shell is

$$\Delta N = 4\pi \varrho r^2 \Delta r$$

and the intensity of light received at the origin from the shell will be

$$\Delta I = k\varrho \, \Delta r \, ,$$

where k is a luminosity coefficient. Hence integrating we have

$$I = k\varrho \int_0^\infty \mathrm{d}r = \infty \, , \tag{6.1}$$

in performing this integration we have of course assumed 4), since as $r \to \infty$, because of the finite propagation-time of light, $t = r/c$, $t \to \infty$.

Now the fact that we obtain infinity above is of course no great mystery. From our assumption 4) it follows a star must have radiated an infinite amount of energy. If the system is to be in statistical equilibrium it must have also received an infinite amount of energy, which is precisely what the above formula tells us.

Now of course there are many ways out of this difficulty. One way is to introduce a non-uniform density $\varrho = \varrho(r)$ such that

$$\int \varrho(r) \, \mathrm{d}r = \text{finite} \, , \tag{6.2}$$

while still permitting the total number of stars to be infinite (*)

$$4\pi \int \varrho(r) \, r^2 \, \mathrm{d}r = \text{infinite} \, . \tag{6.3}$$

For example setting $\varrho = \varrho_0 \big(1 + (r^2/R^2)\big)^{-1}$, where R is some fundamental cosmological length. The difficulty with this model is of course the problem of conservation of energy. After all how can a star continue to radiate forever with a constant luminosity? We can calculate life times for stars and they are of the order of $\sim 10^{10}$ years (this of course was not known during Olbers time).

Let us now assume that the stars have a finite lifetime T and for simplicity they radiate with constant luminosity during that period. Then it is clear that we can dispose of the infinity in (6.1), by the following assumption:

(*) This is a simple example of the so-called hierarchical models.

421

4') The stars were all created simultaneously (with a uniform density throughout all of space) at a time $t_0 < T$ in the past.

With 4') the total intensity received will be

$$I = k\varrho c t_0 \, . \tag{6.4}$$

There are of course many difficulties with this hypothesis from the standpoint of our usual physical ideas, nevertheless as we shall now see all other present cosmologies encounter similar difficulties either at a singular point in space at an initial time, or at singular points throughout all space for all times. In any case if we make rough estimates for the above quantities k, ϱ, and also using the fact that the intensity of starlight is $\approx 10^{-3}\,\text{erg/cm}^2\cdot\text{s}$ we find $t_0 \approx (10^8 \div 10^{10})$ years in agreement with above. As we shall see there are other ways of arriving at this magnitude of time.

6·4. – Various cosmological models.

The possibility that space is infinite and that there is an infinite number of stars created at a finite time t_0 in the past is of course unattractive because of the singling out of time over space and also the difficulties in imagining how such a strange event could have ever occurred. One might therefore imagine a model in which space is finite (for example spherical) and static such as was envisaged in an early model by Einstein.

We imagine the stars or nebulae distributed throughout space, just as dots on the surface of a 3-sphere. We therefore have a finite number of stars. However even this is not sufficient to guarantee the low brightness of the night sky, unless the stars have been losing energy effectively for only a finite period, since otherwise in such an « enclosure », the radiation content would have continued to increase until the background was as bright as the stars.

The difficulty with this model is similar to that of the infinite distribution model (created at time t_0 in the past), except that one has a finite number created simultaneously at t_0 distributed throughout a finite space. This singling out of time as a singular point in preference to space however is unsatisfactory from the standpoint of relativity and one would prefer to find space itself also become singular when time did. Proceeding in this way we arrive at the so-called expanding (or explosion theory) models of the universe. In these models we imagine matter or energy to have been highly concentrated at some time in the past. Since from the standpoint of Einstein's field equations the distribution of energy to some extent determines the structure of space itself (*),

(*) As we have remarked before it does not do this entirely, *e.g.* even if $T_{\mu\nu}=0$, we have many solutions possible to $G_{\mu\nu}=0$; flat space, hyperbolic space, etc.

422

we may therefore look upon this singular point as in some sense the beginning of space, time and matter. After some time interval Δt, the system will have spatial volume ΔV, and in energy density ρ so that thereafter we can apply our physical laws.

There are (at least) two objections against this model; one, physical, the other, philosophical. The physical objection concerns itself with the problem of anti-matter. We know from experiment that there are anti-protons, anti-neutrons, positron s, etc., and the question is, « what became of these anti-particles? »

Part of the reason for the unsolved character of this difficulty is that it was not contemplated when these models were first formulated (1920's). Indeed, even in the early 1930's, people thought that the electron was in some sense the anti-proton, until it was shown both experimentally and theoretically that this could not be the case. When the question was finally raised in connection with cosmology no one could give a satisfactory answer, beyond speculating that perhaps the antimatter was in another part of the universe (for example on the other side of the « sphere », if the universe is spherical). Although no satisfactory explanation has ever been given as to how it got there, *i.e.*, how the two kinds of matter become separated. One might speculate that this has something to do with charge conjugation and parity violations; but it is still quite an open question. Before going on to the philosphical difficulty it might be well to call attention to an early difficulty which no longer exists due to a change in the experimental data. In the 1930's HUBBLE estimated from observing the red shift associated with the distant nebulae, an « age » of the universe of about 10^9 years. (The approximate formula is

$$\text{red shift} = (\text{Hubble's costant}) \times \frac{\text{distance}}{c}$$

$$\frac{\delta\lambda}{\lambda} = H\frac{r}{c}\,.$$

The reciprocal of Hubble's constant, H^{-1} is therefore a measure of the age of the universe if the red shift is interpreted as a Doppler effect due to an expansion process. We shall discuss various relativistic models in greater detail anon.)

On the other hand the age of the earth as determined from radioactive decay processes turned out to be of order $5 \cdot 10^9$ years, and a similar figure from meteorites. This discrepancy persisted for over 20 years (1934-1955) and led to a great many erroneous conclusions, in particular that relativistic cosmology was in disagreement with experiment. However recent measurements of SANDAGE (*) and others would tend to set the value of Hubble's age parameter

(*) *Astrophys. Journ.*, **127**, 513 (1958).

423

at $> 6 \cdot 10^9$ years and even as much as $\sim 2 \cdot 10^{10}$ years, so that there is no longer a difficulty (*).

We therefore turn to the philosophical difficulty. Various people such as BONDI and GOLD, and particularly F. HOYLE have argued that it is philosophically more satisfying if the universe did not start from a singularity in space and time, but actually is « continuously created », in accordance with the so-called « Perfect Cosmological Principle ». According to this principle (or really, hypothesis), the universe presents essentially the same aspect throughout all time and throughout all space. That is to say, the universe is stationary. The importance of this stationary or steady state property, is that there is no unique point in time singled out as « the beginning », which could never be explored by means of experiment. Now, as we have seen from our discussion of Olbers paradox, we encounter difficulties if we have a universe infinite in space and time with a uniform distribution of matter (which must be the case according to the perfect cosmological principle). The way out of these difficulties according to the steady-state hypothesis is to assume, (in agreement with observation) that the universe is expanding, or to put it differently, the distances between nebulae are increasing with time. Now because of the Doppler shift, the light coming from the more distant nebulae will be shifted towards the red, and if we assume this shift is just proportional to distance, then this will have the result of a finite intensity of light. Indeed in the steady-state model the line element is taken to be

$$ds^2 = dt^2 - \exp[2t/T](dx^2 + dy^2 + dz^2) ,$$

where T is the reciprocal of Hubble's constant, and is a measure of the age of the universe (in non steady-state models). One can show setting $ds = 0$, that the red shift for light of wavelength λ coming from a nebula of distance r, will be

$$\frac{\delta\lambda}{\lambda} = \frac{r}{T} .$$

Now of course if the distance between nebulae increases with time, the density of nebulae will diminish and this is not in agreement with the perfect cosmological principle. They therefore are led to assume that matter is constantly appearing in interstellar space so as to preserve a constant density. It is for this reason one sometimes refers to this model as the « continuous creation » model. It is to be emphasized that there is no direct experimental evidence in favor of this hypothesis. On the other hand, there are grave difficulties with micro-physics. Thus HOYLE assumes that hydrogen atoms are continually ap-

(*) See however, A. SANDAGE: *Ap. J.*, **133**, 355 (1961). There is now a discrepancy with stellar age estimates.

421

pearing in interstellar spaces, which then condense to form stars and nebulae, the nebulae gradually separating and new hydrogen appearing.

One of the major difficulties with this hypothesis is that it violates our usual microphysical conservation laws: conservation of baryons, conservation of leptons, local conservation of energy, angular momentum, etc.

The fact that the rate of production of hydrogen according to HOYLE is small

$$\left[\frac{\Delta m}{\Delta t\,\Delta^3 x} \approx \frac{3\varrho_0}{T} \approx 10^{-16}\,\mathrm{g/cm^3\,s}\right],$$

is irrelevant, since the above conservation laws are supposed to be « strict » conservation laws. Indeed about the only thing one isn't giving up in such a model is the conservation of charge and possibly angular momentum if one assumes the hydrogen is « created » in an s-state with the spins of the proton and electron oppositely aligned. One might try to conserve baryon and lepton number by assuming the simultaneous production of neutrons and anti-neutrons but then it would be difficult to understand why the anti-neutrons and neutrons do not annihilate with one another.

Thus although the steady state model disposes of Olbers paradox and presumably satisfies the perfect cosmological principle, it gives rise to a great many difficulties with respect to microscopic conservation laws. Since the number of things one has been able to predict and explain using these laws is so enormously greater than what one has been able to measure and predict according to the steady-state hypothesis, we cannot help but feel some hesitation towards accepting it. However within the next decade the observation of nebulae red shift, nebulae count, density of matter in the universe, stellar populations, relative abundance of the elements, etc., together with further developments of our knowledge of elementary particle physics should enable us to decide one way or the other on this bold hypothesis.

6·5. – Summary of general relativistic cosmological models.

In the preceding sections we have tried to outline some of the principal problems in formulating cosmological models. As we have seen they are rather formidable when we do not ignore the fundamental questions that must be asked and answered if such models are to have any profound significance. Let us now, however, for pedagogical purposes, ignore some of these more difficult questions and attempt to summarize the mathematical scheme used in general relativistic cosmologies.

First, what constitutes the primary observational data? As we have remarked before, a study of distant galaxies reveals a shift in the spectral lines $\Delta\lambda/\lambda$ which increases very nearly linearly with distance and is apparently

isotropic. Needless to say, when one says « proportional to distance », the question arises as to just how one measures distance, since we have no direct way of doing this. The present method consists in making the assumption that the absolute luminosity of an average nebula is always the same (*). There is a wealth of problems connected with obtaining reliable distance measures (*e.g.* for near nebulae using the period luminosity relation for cepheid variable stars) but from our standpoint the most important theoretical point is to note that the method essentially involves the assumption that the universe is homogeneous, otherwise we should have no way of being able to infer from the photographic magnitudes of the distant galaxies, their absolute magnitudes and hence distance.

We therefore commence to build our general relativistic cosmology on the following hypotheses:

1) Space is isotropic.

2) Space is homogeneous.

3) The field equations of general relativity are valid for determining the metric.

However these hypotheses are obviously inadequate, since

a) There are two forms of the field equations:

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = - \varkappa T_{\mu\nu}\,,$$

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = - \varkappa T_{\mu\nu} - \Lambda g_{\mu\nu}\,.$$

b) We don't know what to insert for $T_{\mu\nu}$, and in the second case what value to choose for Λ.

To proceed further therefore we make some simplifying hypotheses concerning $T_{\mu\nu}$: namely, that we can pretend the galaxies have been smeared out so that they form a « dust » with an energy momentum tensor of the form

$$T^{\mu}_{\nu} = (\varrho_0 + p)\,\frac{\mathrm{d}x^{\mu}}{\mathrm{d}s}\,\frac{\mathrm{d}x_{\nu}}{\mathrm{d}s} - \delta^{\mu}_{\nu} p\,,$$

where p is a scalar pressure and ϱ_0 a density of energy (or mass). Sometimes the energy momentum tensor for the electromagnetic field is added, although

(*) Apart from the so-called Stebbins-Whitford reddening which would imply a dependence of luminosity on time (which incidentally is excluded in the steady-state theories) and indeed an additional redding as we progress backwards in time to more distant nebulae. Recent observations tend to make this additional shift very small, if it is present at all.

its contribution is presumed small since the mean density of the « dust » is of order $(10^{-30} \div 10^{-29})$ g/cm^3, whereas the comparable density of radiant energy is considerably smaller and therefore can be neglected in this approximate type of calculation. To make things even simpler one goes further and assumes that the kinetic energy of the dust is small so that all one has left is ϱ, p so that terms of the form

$$T^i_0 = T^0_i = T^i_j = 0 \qquad \begin{pmatrix} i, j = 1, 2, 3 \\ i \neq j \end{pmatrix}.$$

One chooses a simple co-ordinate system to express the line element using hypotheses 1) and 2) and the following assumptions:

4) Clocks at rest with respect to the dust keep time at the same rate independently of position and elapsed time so that there is effectively an absolute time read by all clocks to within a synchronization of the zero point.

5) The time (measured with such clocks) for light to propagate from a point A of the dust to a point B is the same as from B to A, if the zero point of time is taken to be the same.

Remark: note that these two assumptions are in accordance with the homogeneity and isotropy property of space extended to the time domain. Indeed when the astronomers assume the absolute magnitude of average nebulae to be the same, since the light from these nebulae is coming from the past as well as distant points in space, one is effectively assuming the validity of 4). Needless to say, one might make other assumptions such as that the rate of a clock in the past was slower than the same clock now and attempt to interpret the red shift this way. But we shall not consider such approaches here as they entail a great deal of complications.

It follows from 4) and 5) that the line element can be written in the form (*)

$$\mathrm{d}s^2 = \mathrm{d}t^2 - g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,.$$

When we do this we are using the so-called « co-moving » co-ordinates which may be visualized in the following way: Imagine for simplicity space to be closed and the nebulae on the surface of a 3-sphere with radius $R(t)$, where $R(t)$ is a single-valued function of time which we shall assume to be increasing.

(*) That the line element can be always written in this form follows from general covariance and mathematics, but it is important to clarify the physical reasons why we choose to express the line element in this form and work in this co-ordinate system.

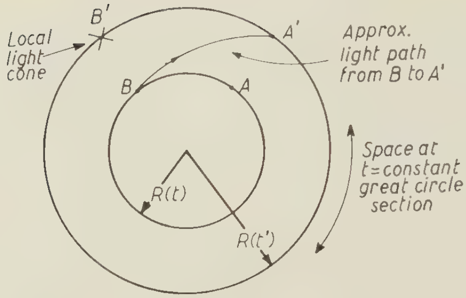


Fig. 8.

Consider now two nebulae A and B at rest relative to the surface of the sphere with radius $R(t)$, then this will hold for all times even though they become more and more separated. Thus the « spatial part » of the line element will be of the form

$$d\Sigma^2 = R(t)^2\,d\sigma^2\,,$$

where $d\sigma^2$ does not depend on time but on the geometry of the space. Also by the above assumptions we have

$$ds^2 = dt^2 - d\Sigma^2 = dt^2 - R(t)^2\,d\sigma^2\,, \tag{6.5}$$

so that the proper time of a clock at rest relative to the surface of radius $R(t)$ is simply the co-ordinate time. We see also that terms of the form $g_{0i}\,dt\,dx^i$ do not appear because of the symmetry in the above picture. Also the fact that the proper time equals the coordinate time is clear, for suppose $g_{00}=g_{00}(t, x^i)$. Now because of the rotational invariance it could only depend on the angle between the lines OA, OB, so $g_{00}=g_{00}(\theta, t)$, however what would we choose for θ_0? There is no preferred nebula. Hence, we set $g_{00}=g_{00}(t)$. But this implies that the rate of a clock depends on time, in disagreement with (4), hence we integrate to obtain a time $t=\int\sqrt{g_{00}(t')}\,dt'$, which we call « cosmic time ». Note that this cosmic time is in some sense a Newtonian concept since we are saying time flows uniformly for observers at rest with respect to the sphere of galaxies. However this idea is fundamental to special relativity: two clocks at rest with respect to a given Lorentz frame keep time (uniformly) at the same rate. Finally, for simplicity in the above discussion we have assumed a closed spherical space, but a hyperbolic or Euclidean space could also have been used as well.

Indeed it has been shown by the mathematicians (Schur's theorem) that the only homogeneous isotropic spaces are those of constant curvature with the curvature positive (spherical or elliptic space), zero (Euclidean or « parabolic »), negative (hyperbolic).

The spatial line element for such spaces may always be written (in isotropic form)

$$\mathrm{d}\sigma^2 = \frac{(\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2}{[1 + (k/4)((x^1)^2 + (x^2)^2 + (x^3)^2)]^2},$$

and hence the total line element is of the form

$$\mathrm{d}s^2 = \mathrm{d}t^2 - R(t)^2 \left[\frac{\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2)}{[1 + (k/4)r^2]^2}\right].$$

For $k = +1$, we have spherical space, $k = 0$, Euclidean, $k = -1$, hyperbolic. We see therefore from the standpoint of the line element there are two quantities to determine:

a) the curvature of the space k,

b) the function $R(t)$.

We now seek to relate these quantities to the energy-momentum tensor which we take in accordance with our previous assumptions. Substituting the above values for $g_{\mu\nu}$ in the field equations we find:

$$3\,\frac{k + \dot{R}^2}{c^2 R^2} = \Lambda + \varkappa \varrho(t),$$

$$2\,\frac{R\ddot{R} + \dot{R}^2 + k}{c^2 R^2} = \Lambda - \frac{\varkappa p(t)}{c^2}.$$

We therefore arrive at a set of simultaneous equations to solve. Unfortunately to determine $R(t)$ we need to know the four quantities k, Λ, $\varrho(t)$, $p(t)$, none of which are known! A rigorous solution is therefore impossible. All we can do is to obtain solutions for a variety of cases and try to check these as well as we can with observations.

The first simplifying assumption is that the pressure term is negligible. This would not be true for radiation since $p = \frac{1}{3}\varrho c^2$ but the radiation density is estimated to be much smaller ($\sim 10^{-5}$) than the smeared out matter density, so this is no difficulty at least at this stage of development. For highly condensed states, at earlier periods, this assumption may be invalid. The covariant energy-conservation law is

$$\frac{\partial\sqrt{-g}}{\partial x^\mu}\, T^\mu_\nu - \frac{1}{2}\frac{\partial g_{\alpha\varrho}}{\partial x^\nu}\, T^{\alpha\varrho}\sqrt{-g} = 0 = \sqrt{-g}\, T^\mu_{\nu;\mu}.$$

Since we have only one component for T^μ_ν, namely T^0_0, because of the above

429

form for the metric, it follows that this conservation law reduces to

$$\frac{\partial}{\partial t}\left(R(t)^3 \varrho(t)\right) = 0 \,,$$

or

$$R(t)^3 \varrho(t) = \text{const} \,.$$

Following standard terminology we define this constant so that

$$\tfrac{1}{3}\varkappa \varrho(t) R(t)^3 = E \,,$$

where E is a measure of the total energy of the system (in suitable units), and is a positive quantity.

Then the above set of equations may be rewritten as (*)

$$\dot{R}^2 = \frac{E}{R} - k + \frac{c^2 \Lambda}{3} R^2 \,.$$

We now consider various cases corresponding to the three values of k, different values of Λ (relative to E). As can be seen on the next page there are a large variety of cases from which to choose, depending on one's philosophical predilections and interpretation of the data. Thus for all three cases ($k = +1$, 0, -1) there are the so-called oscillating models. At some time t_0 in the past space collapses $\left(R(t) = 0\right)$ and the density $\varrho(t)$ becomes infinite. In many respects this is very difficult to visualize if $k = 0$, -1, since we don't have a closed space. On the other hand, for a closed space $k = 1$, such a condensation presents no difficulty. Other cases for sufficiently large Λ correspond to monotonically expanding universes. There is no agreement on the data at the present although the model of Hoyle—flat, ever expanding—appears unlikely.

The double line separates the oscillatory models from the continuously expanding ones.

Historically, the first general relativistic model proposed (1917) was a static one, Einstein's universe with $\dot{R} = 0$, $k = 1$. It was in this model that the cosmological term was first introduced, with the hope that it would also serve to exclude solutions for $T_{\mu\nu} = 0$ in conformity with Mach's principle. This hope was shown to be unjustified by de Sitter shortly afterwards, who obtained the solution, corresponding in this co-moving co-ordinate system (G. Lemaître,

(*) The second equation may be obtained by differentiation of the above equation with $p(t) = 0$.

430

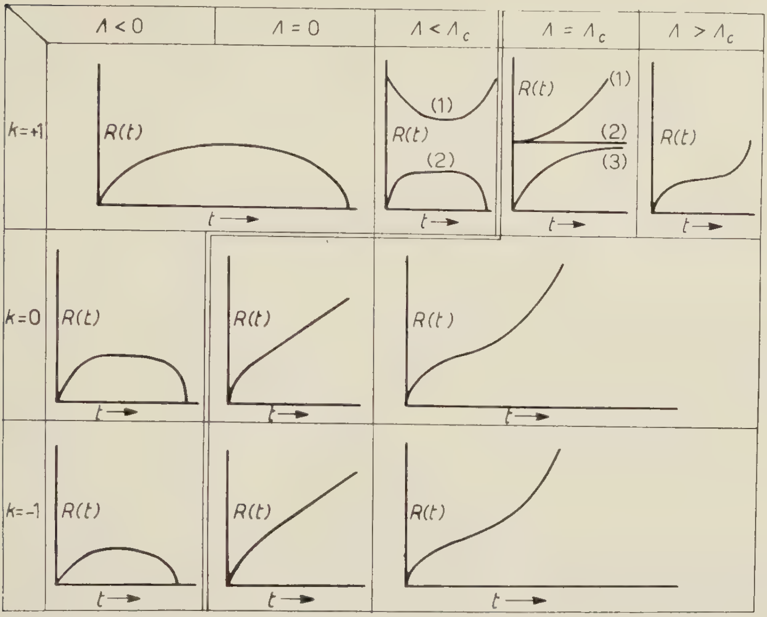


Fig. 9. – Behavior of $R(t)$ for various model universes.

H. P. ROBERTSON) to $E = 0$ (*i.e.* no matter present) and $k = 0$, with $R(t) = \exp[t/T]$, $T = \sqrt{3/\Lambda}$. The situation was greatly clarified by the wokr of FRIEDMANN (1922). The fact that one could obtain closed universes without the cosmological term, while maintaining a reasonable structure for the $T_{\mu\nu}$ led EINSTEIN to later conclude (1931) that the cosmological term should be discarded—on the basis of simplicity. The difficulty then remains that if one sets $T_{\mu\nu} = 0$, the field equations admit the Minkowski line element as a valid solution, which, as indicated previously, is very puzzling from the stand point of Mach's principle.

6·6. – Calculation of red-shift from line element for model universe.

Let us now see how we may obtain a red shift versus distance relation from the above line element and how it may be used to decide on the preceding possibilities. First of all we note that the above isotropic form may be written as

$$\text{(6.6)} \qquad \mathrm{d}s^2 = \mathrm{d}t^2 - R(t)^2 \left[\frac{\mathrm{d}r'^2}{1 - kr'^2} + r'^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\right],$$

upon setting

$$\text{(6.7)} \qquad r' = \frac{r}{1 + (k/4)r^2},$$

in the preceding form. The above form is much more convenient for our purposes, since we shall consider the light rays from the distant galaxies as propagating along the lines $d\theta = 0$, $d\varphi = 0$. The quantity u defined as (dropping the prime on r for convenience)

$$u = \int_0^r \frac{dr}{(1 - kr^2)^{\frac{1}{2}}}, \tag{6.8}$$

is the proper spatial distance between our galaxy, say, and a galaxy at a coordinate distance r away.

Note that for kr^2 sufficiently small we may expand the integral to obtain

$$u = r + \frac{k}{6} r^3 + \dots . \tag{6.9}$$

Thus it is necessary to go to third order to obtain a difference between the three cases ($k = 1,\ 0,\ -1$).

We now set $ds = 0$, to find the time necessary for light, emitted from a galaxy at time t_1 in the past, to propagate to our own: arriving at t_0. We have

$$\int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_0^r \frac{dr}{(1 - kr^2)^{\frac{1}{2}}} = u . \tag{6.10}$$

Now, holding the proper distance fixed, consider the variation in the integral for light emitted at time $t_1 + \Delta t_1$ to be received at time $t_0 + \Delta t_0$ (corresponding to the emission time Δt_1 at the distant galaxy for light of a given wave length) we have

$$\Delta \int_{t_1}^{t_0} \frac{dt}{R(t)} = \frac{\Delta t_0}{R(t_0)} - \frac{\Delta t_1}{R(t_1)} = 0 . \tag{6.11}$$

Now since the proper time of the clocks corresponds to their co-ordinate time $\Delta t_1 \propto$ wave length of light emitted, $\Delta t_0 \propto$ to the wavelength received, hence

$$\frac{\lambda_0}{\lambda_1} = \frac{R(t_0)}{R(t_1)} . \tag{6.12}$$

Introducing the notation $z = \Delta\lambda/\lambda$ we have

$$z = \frac{R_0}{R_1} - 1, \tag{6.13}$$

432

We shall assume $R(t_1)$ may be expanded

$$R(t_1) = R_0 - t\dot{R}_0 + \frac{t^2}{2}\ddot{R}_0 + \dots , \tag{6.14}$$

and use this relation together with (6.10) to express $Z = f(u, R_0, \dot{R}_0, \ddot{R}_0)$ which will then relate the red shift to proper distance and R_0, $\dot{R}_0$, $\ddot{R}_0$. We have

$$u = \int_{-t}^{0} \frac{dt'}{R_0 + t'\dot{R}_0} \quad \text{or} \quad t = uR_0 - \frac{1}{2}u^2 R_0 \dot{R}_0 , \tag{6.15}$$

substituting this value for t in (6.14) and the result in (6.13) we find

$$z = \dot{R}_0 u + \tfrac{1}{2}(\dot{R}_0^2 - R_0\ddot{R}_0)u^2 + \dots . \tag{6.16}$$

It is convenient here to restore dimensions using (6.15), then we have $r \approx cR_0 u$, locally, hence

$$z = \frac{\dot{R}_0}{R_0}\frac{r}{c} + \frac{1}{2c^2}\left(\frac{\dot{R}_0^2 - R_0\ddot{R}_0}{R_0^2}\right) r^2 + \dots , \tag{6.17}$$

neglecting terms of order (r^3). The quantity $R_0/\dot{R}_0$ is called Hubble's constant H, introduced previously.

Experimentally one of course does not measure r directly, but as stated previously, galactic apparent luminosity l (energy per unit area per unit time) which, measured on a logarithmic scale is its apparent magnitude m. To arrive at an expression for magnitude m, it is convenient to consider the light emitted in the form of photons.

Then we have for the number of photons per unit area N_1 emitted at time t_1 in the interval δt_1, with energy hc/λ_1, and the number per unit area N_0 received at time t_0, in the interval δt_0, with energy hc/λ_0,

$$\begin{cases} \text{emitted} = N_1 \dfrac{hc}{\lambda_1}\dfrac{1}{\delta t_1} , \\ \text{received} = N_0 \dfrac{hc}{\lambda_0}\dfrac{1}{\delta t_0} . \end{cases} \tag{6.18}$$

But we have $\lambda_0 = \lambda_1(1+z)$, $\delta t_0 = \delta t_1(1+z)$ and clearly

$$N_0 = \frac{4\pi r_1^2 R(t_1)^2 c^2 N_1}{4\pi r_0^2 R(t_0)^2 c^2}$$

hence l is given by

$$l_0 = \frac{L_1}{4\pi r^2 R(t_0)^2 (1+z)^2 c^2}, \tag{6.19}$$

where $L_1 = l_1 4\pi r_1^2 R(t_1)^2 c^2 N_1(hc/\lambda_1 \delta t_1)$ (energy/time) and is taken to be the same for all the galaxies (apart from the Stebbins-Whitford effect) ont he average. The magnitude is then expressed by

$$m = \text{const} - 2.5 \log_{10} l_0, \qquad m = \text{const}' - 5 \log_{10}(1+z)r. \tag{6.20}$$

Substituting for r one obtains a formula given *e.g.* by ROBERTSON (*)

$$m = M_0 - 45.06 + 5 \log z/H + 1.086 \left(1 + \frac{\ddot{R}_0}{R_0 H^2} - 2\mu\right) z + O(z^2), \tag{6.21}$$

where M_0 is the absolute magnitude and the term -2μ has been added which allows for changes in magnitude, absorbtion of light in interstellar spaces. The quantity $-\ddot{R}_0/R_0 H^2$ is the so-called deceleration parameter q. If q were large and positive we could say with confidence the expansion is slowing down. The work of SANDAGE would imply $q = 2.6 \pm .8$. However Baum (**) most recently found $q \approx 5$ to 1.5. On the other hand Hoyle's model yields $q = -1$ and this is ruled out both by data of BAUM and SANDAGE.

Without a good knowledge of q it is not possible to select from the various models with any certainty although of course the Einstein static model is ruled out, and probably the steady-state model as well.

One curious possibility, frequently considered, is that the universe is closed and *oscillatory*. Under these circumstances, eventually the nearby galaxies and the distant galaxies would exhibit a blue shift. Ultimately there would be a reuniting of the galaxies into the initial singular state. (It is to be emphasized that the singular state need not be a true singularity but actually a small space-time region where our approximations break down.) Such a universe in a certain sense satisfies the perfect cosmological principle since, on the average of many oscillations, it would present the same appearance throughout time and, at a given time, would be homogeneous and isotropic in space.

However a basic criticism that can be levelled against this model as well

---

(*) *Jubilee of Relativity Theory*, in *Helv. Phys. Acta, Supplementum* IV (Basel, 1956), p. 128.
(**) *Astron. Journ.*, **62**, 6 (1957).

as the steady-state model, and indeed all proposed models, is that they are in a sense too mechanistic. On the other hand, the novel phenomena associated with quantum mechanics and life itself, suggest that the structure of the universe has a richer and more subtle character than one would infer from these highly simplified, albeit ingenious models. Thus the future development of the theory must find a unified way of relating cosmology to quantum physics.

Chapter VII.

SUPPLEMENTARY REMARKS.

For brevity we have omitted the following topics which would form the basis for a more advanced course:

1) Reissner-Nordström solution for a charged particle; structure of the energy-momentum tensor of a point particle.
2) Systems with axial symmetry.
3) Variational principle for the field equations.
4) Definition of an energy-momentum complex for the gravitational field.
5) Gravitational radiation theory.
6) Derivation of the Newtonian and « post-Newtonian » equations of motion of concentrated sources from the field equations using perturbation theory.
7) Non-Riemannian geometries: spaces with torsion, Finsler space.
8) Attempts at a unification of gravitation with electro-magnetism.
9) Attempts at quantizing general relativity.
10) Topological considerations.

It is perhaps interesting to remark in connection with 6) and 9) that when Einstein published the first of these investigations (Einstein and Grommer, 1927), he was driven in part by a desire to see whether the field equations might exhibit or lead to quantum conditions. So far this hope has proved unjustified, but it has become increasingly clear in the past several years that unless one has such a deeper basis for unifying the two disciplines, one is unlikely to succeed.

A basic difficulty connected with 5) is that although the weak-field gravitational wave equations have been known for nearly half a century, no such waves have ever been observed. Moreover a simple calculation indicates that their intensity ought to be fantastically small. Because of this, the subject

takes on a very speculative aspect, which becomes all the more so when one sets about attempting to quantize these unobserved waves. In addition to the mathematical complexity of the field equations, the situation is further complicated by the problems connected with 4): one does not have a true energy-momentum tensor for the gravitational field but a complex or « pseudo-tensor » which in general can be made to vanish locally or along a geodesic curve, or in some cases, (*e.g.* Schwarzschild field in quasi-rectangular proper co-ordinates) vanishes everywhere (SCHRÖDINGER, 1918). This is of course totally unlike electromagnetism and implies something very peculiar about the concept of « gravitational energy ».

We defer a more complete discussion of these questions to another occasion.

* * *

The author wishes to thank Prof. E. R. CAIANIELLO for his warm and stimulating hospitality. He also wishes to thank the National Science Foundation and the Office of Naval Research for their partial support.

BIBLIOGRAPHY

P. G. BERGMANN: *Introduction to the Theory of Relativity* (New York, 1947).
H. BONDI: *Cosmology* (Cambridge, 1960).
A. S. EDDINGTON: *Mathematical Theory of Relativity* (Cambridge, 1924).
A. EINSTEIN: *The Meaning of Relativity* (Princeton, 1955).
A. EINSTEIN, H. A. LORENTZ, H. MINKOWSKI and H. WEYL: *The Principle of Relativity* (London, 1923).
A. EINSTEIN: *Essays in Science* (New York, 1953).
V. FOCK: *Theory of Space, Time and Gravitation* (New York, 1959).
Helv. Phys. Acta: *Jubilee of Relativity Theory*, Suppl. IV (Basel, 1956).
L. LANDAU and E. LIFSHITZ: *The Classical Theory of Fields* (Cambridge, Mass., 1951).
A. LICHNEROWICZ: *Théories relativistes de la gravitation* (Paris, 1955).
G. C. MCVITTIE: *General Relativity and Cosmology* (London, 1956).
C. MØLLER: *The Theory of Relativity* (Oxford, 1952).
W. PAULI: *Theory of Relativity* (London, 1958).
G. RAINICH: *Mathematics of Relativity* (New York, 1950).
D. SCIAMA: *The Unity of the Universe* (New York, 1959).
L. SILBERSTEIN: *The Theory of Relativity* (London, 1924).
J. L. SYNGE: *Relativity, The General Theory* (Amsterdam, 1960).
R. C. TOLMAN: *Relativity, Thermodynamics and Cosmology* (Oxford, 1934).

436

PROPRIETÀ LETTERARIA RISERVATA

Direttore responsabile: G. POLVANI Tipografia Compositori - Bologna

Questo fascicolo del *Supplemento* è stato licenziato dai torchi il 15-IV-1961